

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 07:24:19 ; Search time 37 seconds
(without alignments)
1310.898 Million cell updates/sec

Title: US-09-619-032-4
Perfect score: 1880
Sequence: 1 MRALVFHGNLQYAEIPKSEI.....RRLDFAFRAYNDWRGNGEP 364

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1874	99.7	364	18	AAW34643
2	645	34.3	128	23	AAO21657
3	142.5	7.6	649	15	AA47504
4	141.5	7.5	655	22	AA96088
5	126	6.7	1362	22	AA96091
6	122.5	6.5	653	19	AAW54870
7	103	5.5	571	21	AAG16038
8	103	5.5	571	21	AAG48655
9	103	5.5	704	21	AAG16037
10	103	5.5	704	21	AAG48654
					Thermostable alpha AEDII12RA alpha-ga
					Pyrococcus furiosu
					Putative P. abyssi
					Putative P. abyssi
					Super heat resista
					Arabidopsis thalia
					Arabidopsis thalia
					Arabidopsis thalia

11	103	5.5	726	21	AA616036	Arabidopsis thalia
12	103	5.5	726	21	AA648653	Arabidopsis thalia
13	103	5.5	1073	22	AAU36768	Staphylococcus aur
14	103	5.5	1147	22	AAU33972	Staphylococcus aur
15	101.5	5.4	360	22	AA62528	S. epidermidis ope
16	101.5	5.4	376	22	AA62506	S. epidermidis ope
17	101.5	5.4	936	22	ABG18270	Novel human diagno
18	101	5.4	329	17	AA94013	Heat resistant mal
19	100	5.3	929	22	AB70857	Drosophila melanog
20	94.5	5.0	367	23	ABP38833	Staphylococcus epi
21	93.5	5.0	1320	23	AA47603	Drosophila cell cy
22	93	4.9	216	21	AA629078	Arabidopsis thalia
23	93	4.9	218	21	AA629077	Arabidopsis thalia
24	93	4.9	235	21	AA629076	Arabidopsis thalia
25	93	4.9	4829	22	AA697833	Human apoptosis-in
26	92.5	4.9	392	23	AB54675	Lactococcus lactis
27	92.5	4.9	625	22	AB61213	Drosophila melanog
28	92	4.9	375	22	AA62373	S. epidermidis ope
29	91.5	4.9	480	21	AA616318	Pinus radiata UGP
30	91.5	4.9	1009	23	AAO20949	Protein of the Bpm
31	91	4.8	364	21	AA614039	Arabidopsis thalia
32	91	4.8	364	23	AB92380	Herbicidally activ
33	91	4.8	389	21	AA614038	Arabidopsis thalia
34	91	4.8	1248	22	AB61465	Drosophila melanog
35	91	4.8	1417	20	AA31947	Plasmodium falcipa
36	90.5	4.8	317	22	AA696470	Putative P. abyssi
37	90.5	4.8	406	23	ABP38725	Staphylococcus epi
38	89.5	4.8	1668	22	AAU34158	Staphylococcus aur
39	89.5	4.8	2397	22	AAU36672	Staphylococcus aur
40	89.5	4.8	2766	22	AA63972	Amino acid sequenc
41	89	4.7	292	23	AB54281	Lactococcus lactis
42	89	4.7	1039	22	ABG15145	Novel human diagno
43	89	4.7	1073	22	AA69611	C glutamicum prote
44	88.5	4.7	473	21	AA607943	A uridine diphosph
45	88.5	4.7	473	21	AA634044	Zea mays protein f

ALIGNMENTS

RESULT 1

AAW34643

ID AAW34643 standard; Protein; 364 AA.

XX AAW34643;

XX AC (first entry)

DT 27-MAR-1998

DE Thermostable alpha-galactosidase AEDII12RA-alpha-gal-18GC.

XX Alpha-galactosidase; alpha-glycosidase; thermostable enzyme;

KW food processing; alpha glycoside hydrolysis; raffinose;

KW stachyose; verbascose; bean; flatulence; AEDII12RA-alpha-gal-18GC.

XX Thermococcus alcaliphilus strain AEDII12RA.

OS WO9732974-A1.

XX 12-SEP-1997.

XX 05-FEB-1997; 97WO-US01452.

XX 08-MAR-1996; 96US-0613220.

XX (RECO-) RECOMBINANT BIOCATALYSIS INC.

XX Murphy D, Reid J, Rudolph MJ;

XX WPI; 1997-470541/43.

DR N-PSDB; AAT93753.

XX Nucleic acid encoding alpha-galactosidase from Thermococcus

PT alcaliphilus - used in food processing to hydrolyse

PT alpha-glycosides, e.g. raffinose

XX

PS Claim 1; Fig 1; 32pp; English.

XX

CC This protein comprises AEDII12RA-alpha-gal-18GC, a claimed

CC thermostable alpha-galactosidase of Thermococcus alcaliphilus

CC AEDII12RA, a bacterium that shows optimum growth at 85 deg C and

CC pH 9.5. Also claimed are: (1) an isolated polynucleotide (see

CC AAT93753) encoding the alpha-galactosidase; (2) a vector containing

CC the polynucleotide or homologous or complementary sequences; (2)

CC host cells containing the vector; (3) a process for producing the

CC alpha-galactosidase in transformed or transfected host cells; an

CC enzyme showing at least 70% identity to alpha-galactosidase and

CC comprising at least 30 amino acid residues of its sequence; and (4)

CC a method for hydrolysing alpha-galactoside bonds using the enzyme.

CC The enzyme can be used to hydrolyse raffinose to sucrose and glucose

CC in sugar beet processing (raffinose inhibits crystallisation of

CC sucrose), and as a digestive aid to hydrolyse raffinose, stachyose

CC and verbascone in beans and other gassy foods.

XX

SQ Sequence 364 AA;

Query Match 99.7%; Score 1874; DB 18; Length 364;

Best Local Similarity 99.5%; Pred. No. 2.3e-180;

Matches 362; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60

Db :|||||

QY 1 LRALVFHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60

Db :|||||

QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSLRVEAQVQDRREVKEELFEVSPKGFWLPELAY 120

Db :|||||

QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSLRVEAQVQDRREVKEELFEVSPKGFWLPELAY 120

Db :|||||

QY 121 DPIIPAILKNDNGEYLFADGEAMLFSAHLSAIPKPIKPLYPHLIKAQREKFRYISYLLG 180

Db :|||||

QY 121 DPIIPAILKNDNGEYLFADGEAMLFSAHLSAIPKPIKPLYPHLIKAQREKFRYISYLLG 180

Db :|||||

QY 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240

Db :|||||

QY 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240

Db :|||||

QY 241 NILLYGTDIEFIGYRDIAGYRMSVEGLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300

Db :|||||

QY 241 NILLYGTDIEFIGYRDIAGYRMSVEGLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300

Db :|||||

QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDRGE 360

Db :|||||

QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDRGE 360

Db :|||||

QY 361 NGEPL 364

Db :|||||

QY 361 NGEPL 364

Db :|||||

RESULT 2

AAO21657

ID AAO21657 standard; Protein; 128 AA.

XX

AC AAO21657;

XX

DT 19-SEP-2002 (first entry)

XX

DE AEDII12RA alpha-galactosidase 18GC protein.

XX

KW Alpha-galactosidase; chromosomal walking; hydrolysis; saccharide;

KW catalyst; Thermococcus alcaliphilus; AEDII12RA alpha-galactosidase 18GC;

KW enzyme.

XX

OS Thermococcus alcaliphilus.

XX

PN US2002045226-A1.

XX

PD 18-APR-2002.

XX

PF 20-JUN-2001; 2001US-0886400.

XX

PR 08-MAR-1996; 96US-0613220.

PR 28-SEP-1999; 99US-0407806.

PR 19-JUL-2000; 2000US-0619032.

XX

PA (MURP/) MURPHY D.

PA (REID/) REID J.

XX

PI Murphy D, Reid J;

XX

DR WPI; 2002-507240/54.

DR N-PSDB; AAL39618.

XX

PT New nucleic acid encoding heat-stable alpha-galactosidase, and related

PT polypeptides and antibodies, useful for hydrolysis of saccharides -

XX

PS Disclosure; Fig 5; 35pp; English.

XX

CC The invention relates to an isolated nucleic acid that comprises a

CC sequence not reproduced, variants with at least 50% identity, and encoded

CC polypeptides with alpha-galactosidase activity. The polynucleotides of

CC the invention can be used as a source of probes and primers for

CC chromosomal walking and for identifying related sequences. The protein of

CC the invention is used to catalyse the hydrolysis of saccharides; to

CC modify libraries of small molecules which are then screened for

CC particular activities and to generate specific antibodies, useful for

CC detection or purification of the protein of the invention and related

CC peptides. This sequence represents the Thermococcus alcaliphilus

CC AEDII12RA alpha-galactosidase 18GC of the invention.

XX

SQ Sequence 128 AA;

Query Match 34.3%; Score 645; DB 23; Length 128;

Best Local Similarity 99.2%; Pred. No. 5.5e-57;

Matches 127; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60

Db :|||||

QY 1 LRALVFHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60

Db :|||||

QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSLRVEAQVQDRREVKEELFEVSPKGFWLPELAY 120

Db :|||||

QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSLRVEAQVQDRREVKEELFEVSPKGFWLPELAY 120

Db :|||||

QY 121 DPIIPAIL 128

Db :|||||

QY 121 DPIIPAIL 128

Db :|||||

RESULT 3

AAO47504

ID AAR47504 standard; protein; 649 AA.

XX

AC AAR47504;

XX

DT 07-JUL-1994 (first entry)

XX

DE Pyrococcus furiosus alpha amylase.

XX

KW Pyrococcus furiosus; alpha amylase; liquefaction; polymers;

KW glucopolymers; thermostable.

XX

OS Pyrococcus furiosus.

XX

PN EP577257-A.

XX

```

05-JAN-1994.
17-MAY-1993; 93EP-0303801.
09-JUN-1992; 92US-0893928.
(UYJO ) UNIV JOHNS HOPKINS.
Anfinsen CB, Laderman K;
WPI; 1994-009532/02.

Purified Pyrococcus furiosus alpha-amylase - used for the
industrial liquefaction of gluco-polymers at high temps.

Claim 2; Figure 9; 41pp; English.

The purified Pyrococcus furiosus alpha amylase can act on substrates
with a low degree of polymerisation. e.g. glucose polymers as short
as maltotriose. The enzyme can be used for efficient industrial
liquefaction of glucopolymers at high temperatures.

SQ Sequence 649 AA;

Query Match 7.6%; Score 142.5; DB 15; Length 649;
Best Local Similarity 22.8%; Pred. NO. 3.6e-05;
Matches 91; Conservative 59; Mismatches 118; Indels 131; Gaps 24;

QY 25 EKAYIPVIETLIKEIP---FGLNITGYTLKFLPKDI----IDLVKGGIASDLIEIGTS 77
   ||| ||| ||| ||| : : : : : ||| : : : : : ||| : : : : :
Db 29 EKCYWPFLETL--EYPMKVAIHTSGPLIEWL-QDIRPEVIDLLRSLVKRGQVEIVVAG 85
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 78 YTHAILPLPLSRVEAQVQRDREVKE--ELFEVSPKGFWLPELAYDPIIPAILKDNQY 135
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 86 FYEPVLASIP---KEDRIEQLRMKEWAKSIGFDARGVWLTERVWQPELVKTLKESGIDY 142
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 136 LFADGEAMLFS AHLNS-----AIKPIKPLYPHLIKAQREKRFRYISYLL 179
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 143 VIVD-DYHFMSAELSKEELYWPYYTGDGGEVIAVFPID-----EK----LRYLI 186
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 180 GLRELKKAIKL---VFEGKVTLKAV--KDIEAVPVVAVNTAVMLGIGRLPLMNPKKVAS 234
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 187 PFRPVDKVLEYLHSLIDGDESKVAVFHDDGEKFGIWPGTVEWVY-----EKG 233
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 235 WI-----EDKDNILLYGTDIE-----FIGYRDIAGYRMSVEGLLEVIDELNSELC 280
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 234 WLREFFDRISDEKINLMYTELEKYKPRGLVYLPASVY-----FEM-----SEWSL 281
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 281 P-----SELKHS-----RELYLRTSSWAPDKSLRIWREDEGN---ARLNMLSYN 322
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 282 PAKQARLFVEFVNELKVGIFEKYRVEVVGGIW---KNF-FYKYPESNYMHKRLMLVSKL 337
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 323 MRGELAFLAENSDARGWEPLPERRLDAFRAIYND--WRG 359
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 338 VRNN-----PEARKYLLRAQCNDAYWHG 360
   : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

RESULT 4
AAB96088
ID AAB96088 standard; Protein; 655 AA.
XX
AC AAB96088;
XX
DT 29-OCT-2001 (first entry)
XX
DE Putative P. abyssi alpha-amylase.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX
PN FR2792651-A1.
XX

```


us-09-619-032-4.rag

Mon Nov '18 08:39:33 2002

RESULT 7
AAG16038
ID AAG16038 standard; Protein; 571 AA.
XX
AC AAG16038;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 16524.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

us-09-619-032-4.rag

Mon Nov '18 08:39:33 2002

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

```

PR 28-OCT-1999;   99US-0161993.
PR 29-OCT-1999;   99US-0162142.

Query Match          5.5%; Score 103; DB 21; Length 571;
Best local Similarity 22.8%; Pred. No. 0.29;
Matches 84; Conservative 45; Mismatches 110; Indels 130; Gaps 20;

QY 29 IPVIETLIKEIPFGLNITGYTLKFLPKDIIDLVKGGIASDL-----IEIIGTSYTHAIL 83
    :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 101 LPILE-MVPEELHF--TVPYALSFMGDPIKMATLGDINQLPFGVKIEKLRLTKTML 156

QY 84 PLL-----PLSRVEAQVORDREYKEELFEVSPKGFWLPDEL 118
    |||      ||::||::||::||::||::||::||::||::||::||::||::||
Db 157 PLLSELGIIIPRETLLWKLLRSGCAYANSRIHA-VQ-----AEVLVLASGKDMMLPSQ 210

QY 119 AYDPIIPAIL-----KONGYEYLFADGEAMLFSAHLSAIAIKPKPLYPHLIKAOREK 170
    ::||      ||||::||::||::||::||::||::||::||::||::||::||::||
Db 211 EEAKRLHGLLKNCVRCFKDNNGHTLLEDSISLL-----TVIKGTG-----KYRRSW 257

QY 171 RFRYSIS-----YLLG--LRRLRAIKLVF-----EGKVTLKAVKDIEAVP---V 209
    |::||      ||::||::||::||::||::||::||::||::||::||::||::||
Db 258 RYDLVSDFLPSPKGEALAYALDEVGLFNNAVGSVFSTMEDGKI----VKGLAGVPDKGP 313

QY 210 WVAVENTAVMLGIGRLPLMPNPKKVASWIEDKNILL-----YGTDIE 250
    :||::||::||::||::||::||::||::||::||::||::||::||
Db 314 VLLVGHYHMLMGLELGMSE-----AFIKEK-NILFRGMAHPVLYSDNDPAKAFDYGDWIK 367

QY 251 FIGYRDIAGYRMSVEGLELVIDELNSELCPLS---ELKHSGRELTYLRTSSWAPDKSLRIW 307
    _||::||      |:::||::||::||::||::||::||::||::||::||::||
Db 368 VFG-----AYPTATNFLKLDLSKSHVLLFPGGAREALHNRGQY-----KLIIW 411

QY 308 REDEGNARL 316
    |::||
Db 412 PEQQEFVRM 420

RESULT 9
ID AAG16037 standard; Protein; 704 AA.
XX AC AAG16037;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 16523.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
```


XX XX

PF	25-FEB-2000;	2000EP-0301439.	
XX	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	28-APR-1999;	99US-0130891.	
PR	30-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	04-MAY-1999;	99US-0132407.	
PR	05-MAY-1999;	99US-0132484.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137222.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	10-JUN-1999;	99US-0138847.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	
PR	16-JUN-1999;	99US-0139453.	
PR	17-JUN-1999;	99US-0139492.	
PR	18-JUN-1999;	99US-0139454.	
PR	18-JUN-1999;	99US-0139455.	
PR	18-JUN-1999;	99US-0139456.	
PR	18-JUN-1999;	99US-0139457.	
PR	18-JUN-1999;	99US-0139458.	
PR	18-JUN-1999;	99US-0139459.	
PR	18-JUN-1999;	99US-0139460.	
PR	18-JUN-1999;	99US-0139461.	
PR	18-JUN-1999;	99US-0139462.	
PR	18-JUN-1999;	99US-0139463.	
PR	18-JUN-1999;	99US-0139750.	
PR	18-JUN-1999;	99US-0139763.	
PR	21-JUN-1999;	99US-0139817.	
PR	22-JUN-1999;	99US-0139899.	
PR	23-JUN-1999;	99US-0140353.	
PR	23-JUN-1999;	99US-0140354.	
PR	24-JUN-1999;	99US-0140695.	
PR	28-JUN-1999;	99US-0140823.	
PR	29-JUN-1999;	99US-0140991.	
PR	30-JUN-1999;	99US-0141287.	
PR	01-JUL-1999;	99US-0141842.	
PR	01-JUL-1999;	99US-0142154.	
PR	02-JUL-1999;	99US-0142055.	
PR	06-JUL-1999;	99US-0142390.	
PR	08-JUL-1999;	99US-0142803.	
PR	09-JUL-1999;	99US-0142920.	
PR	12-JUL-1999;	99US-0142977.	
PR	13-JUL-1999;	99US-0143542.	
PR	14-JUL-1999;	99US-0143624.	
PR	15-JUL-1999;	99US-0144005.	
PR	16-JUL-1999;	99US-0144085.	
PR	16-JUL-1999;	99US-0144086.	
PR	19-JUL-1999;	99US-0144325.	
PR	19-JUL-1999;	99US-0144331.	
PR	19-JUL-1999;	99US-0144332.	
PR	19-JUL-1999;	99US-0144333.	
PR	19-JUL-1999;	99US-0144334.	
PR	19-JUL-1999;	99US-0144335.	
PR	20-JUL-1999;	99US-0144352.	
PR	20-JUL-1999;	99US-0144632.	
PR	20-JUL-1999;	99US-0144884.	
PR	21-JUL-1999;	99US-0144814.	
PR	21-JUL-1999;	99US-0145086.	
PR	21-JUL-1999;	99US-0145088.	
PR	22-JUL-1999;	99US-0145085.	
PR	22-JUL-1999;	99US-0145087.	
PR	22-JUL-1999;	99US-0145089.	
PR	22-JUL-1999;	99US-0145192.	
PR	23-JUL-1999;	99US-0145145.	
PR	23-JUL-1999;	99US-0145218.	
PR	23-JUL-1999;	99US-0145224.	
PR	26-JUL-1999;	99US-0145276.	
PR	27-JUL-1999;	99US-0145913.	
PR	27-JUL-1999;	99US-0145918.	
PR	27-JUL-1999;	99US-0145919.	
PR	28-JUL-1999;	99US-0145951.	
PR	02-AUG-1999;	99US-0146386.	
PR	02-AUG-1999;	99US-0146388.	
PR	02-AUG-1999;	99US-0146389.	
PR	03-AUG-1999;	99US-0147038.	
PR	04-AUG-1999;	99US-0147204.	
PR	04-AUG-1999;	99US-0147302.	
PR	05-AUG-1999;	99US-0147192.	
PR	05-AUG-1999;	99US-0147260.	
PR	06-AUG-1999;	99US-0147303.	
PR	06-AUG-1999;	99US-0147416.	
PR	09-AUG-1999;	99US-0147493.	
PR	09-AUG-1999;	99US-0147935.	
PR	10-AUG-1999;	99US-0148171.	
PR	11-AUG-1999;	99US-0148319.	
PR	12-AUG-1999;	99US-0148341.	
PR	13-AUG-1999;	99US-0148565.	
PR	13-AUG-1999;	99US-0148684.	
PR	16-AUG-1999;	99US-0149368.	
PR	17-AUG-1999;	99US-0149175.	
PR	18-AUG-1999;	99US-0149426.	
PR	20-AUG-1999;	99US-0149722.	
PR	20-AUG-1999;	99US-0149723.	
PR	20-AUG-1999;	99US-0149929.	
PR	23-AUG-1999;	99US-0149902.	
PR	23-AUG-1999;	99US-0149930.	
PR	25-AUG-1999;	99US-0150566.	
PR	26-AUG-1999;	99US-0150884.	
PR	27-AUG-1999;	99US-0151065.	
PR	27-AUG-1999;	99US-0151066.	
PR	27-AUG-1999;	99US-0151080.	
PR	30-AUG-1999;	99US-0151303.	
PR	31-AUG-1999;	99US-0151438.	
PR	01-SEP-1999;	99US-0151930.	
PR	07-SEP-1999;	99US-0152363.	
PR	10-SEP-1999;	99US-0153070.	
PR	13-SEP-1999;	99US-0153758.	
PR	15-SEP-1999;	99US-0154018.	
PR	16-SEP-1999;	99US-0154039.	
PR	20-SEP-1999;	99US-0154779.	
PR	22-SEP-1999;	99US-0155139.	
PR	23-SEP-1999;	99US-0155486.	
PR	24-SEP-1999;	99US-0155659.	

PR	28-SEP-1999;	99US-0156458.	XX	AC	AAG16036;
PR	29-SEP-1999;	99US-0156596.	AC		
PR	04-OCT-1999;	99US-0157117.	XX	DT	17-OCT-2000 (first entry)
PR	05-OCT-1999;	99US-0157753.	XX		
PR	06-OCT-1999;	99US-0157865.	XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 16522.
PR	07-OCT-1999;	99US-0158029.	XX		
PR	08-OCT-1999;	99US-0158232.	XX	KW	Protein identification; signal transduction pathway; metabolic pathway;
PR	12-OCT-1999;	99US-0158369.	KW	KW	hybridisation assay; genetic mapping; gene expression control; promoter;
PR	13-OCT-1999;	99US-0159293.	KW		termination sequence.
PR	13-OCT-1999;	99US-0159294.	XX		
PR	13-OCT-1999;	99US-0159295.	XX	OS	Arabidopsis thaliana.
PR	14-OCT-1999;	99US-0159329.	XX		
PR	14-OCT-1999;	99US-0159330.	XX	PN	EPI033405-A2.
PR	14-OCT-1999;	99US-0159331.	XX		
PR	14-OCT-1999;	99US-0159637.	PD		
PR	14-OCT-1999;	99US-0159638.	XX		
PR	18-OCT-1999;	99US-0159584.	XX		
PR	21-OCT-1999;	99US-0160741.	PF		
PR	21-OCT-1999;	99US-0160767.	XX		
PR	21-OCT-1999;	99US-0160768.	PR		
PR	21-OCT-1999;	99US-0160770.	PR		
PR	21-OCT-1999;	99US-0160814.	PR		
PR	21-OCT-1999;	99US-0160815.	PR		
PR	22-OCT-1999;	99US-0160980.	PR		
PR	22-OCT-1999;	99US-0160981.	PR		
PR	22-OCT-1999;	99US-0160989.	PR		
PR	25-OCT-1999;	99US-0161404.	PR		
PR	25-OCT-1999;	99US-0161405.	PR		
PR	25-OCT-1999;	99US-0161406.	PR		
PR	26-OCT-1999;	99US-0161359.	PR		
PR	26-OCT-1999;	99US-0161360.	PR		
PR	26-OCT-1999;	99US-0161361.	PR		
PR	28-OCT-1999;	99US-0161920.	PR		
PR	28-OCT-1999;	99US-0161992.	PR		
PR	28-OCT-1999;	99US-0161993.	PR		
PR	29-OCT-1999;	99US-0162142.	PR		
Query Match 5.5%; Score 103; DB 21; Length 704;					
Best Local Similarity 22.8%; Pred. No. 0.4;					
Matches 84; Conservative 45; Mismatches 110; Indels 130; Gaps 20;					
QY	29	IPVIETLIKEIPFGLNITGYTLKFLPKDIDLVKGGIASDL-----IEIIGTSYTHAIL 83			
Db	234	LPILE-MVPEELHF---TPYALSFMGDPKMATLGIDNQLPTGVKIEKLRQLTKTML 289			
QY	84	PLL-----PLSRVEAQVQRDREVKEELFEVSPKGFWLPEL 118			
Db	290	PLLSELGGIIPRETLLWKLLRSGCAYANSRIHA-VQ-----AEVLVLASGKDMMLPSQ 343			
QY	119	AYDPIIPAIL-----KNGYEYLFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREK 170			
Db	344	EEAKRLHGLLNKNCVRCFKDNGHTLLLEDLSLL-----TVIKGTG-----KYRRSW 390			
QY	171	RFRYIS-----YLLG--LRELKAIKLVF-----EGKVTLKAVKDIEAVP--V 209			
Db	391	RYDLVSDFLPPSKGELAYALDEVGLFRLNAVGVSVFFSTMEDGKI-----VKGLAGVPDKGP 446			
QY	210	WVAVNTAVMLGIGRLPLMNPKKVASWIEDKNILL-----YGTDIE 250			
Db	447	VLLVGYHMLMGLGELPMSE-----AFIKER-NILFRGMHPVLYSDNDPAKAFDYGDWIK 500			
QY	251	FIGYRDIAGYRMSVEGLEVIDELNSELCLPS---ELKHSGRELYLRTSSWAPDKSLRIW 307			
Db	501	VFG-----AYPVTATNLFKLLDSKSHVLLFPGGAREALHNRGEQY-----KLIW 544			
QY	308	REDEGNARL 316			
Db	545	PEQQEFVRM 553			
RESULT 11					
AAG16036					
ID AAG16036 standard; Protein; 726 AA.					

XX	17-OCT-2000	(first entry)	99US-0121825.
AC			99US-0123180.
XX			99US-0123548.
DT			99US-0125788.
XX			99US-0126264.
DE	Arabidopsis thaliana protein fragment	SEQ ID NO: 16522.	99US-0126785.
XX			99US-0127462.
KW	Protein identification; signal transduction pathway; metabolic pathway;		99US-0128234.
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		99US-0128714.
KW	termination sequence.		99US-0129845.
XX			99US-0130077.
OS	Arabidopsis thaliana.		99US-0130449.
XX			99US-0130510.
PN	EPI033405-A2.		99US-0130891.
XX			99US-0131449.
PD			99US-0132048.
XX			99US-0132407.
XX			99US-0132484.
XX			99US-0132485.
XX			99US-0132486.
XX			99US-0132487.
XX			99US-0132863.
XX			99US-0134256.
XX			99US-0134218.
XX			99US-0134219.
XX			99US-0134221.
XX			99US-0134370.
XX			99US-0134768.
XX			99US-0134941.
XX			99US-0135124.
XX			99US-0135353.
XX			99US-0135629.
XX			99US-0136021.
XX			99US-0136392.
XX			99US-0136782.
XX			99US-0137222.
XX			99US-0137528.
XX			99US-0137502.
XX			99US-0137724.
XX			99US-0138094.
XX			99US-0138540.
XX			99US-0138847.
XX			99US-0139119.
XX			99US-0139452.
XX			99US-0139453.
XX			99US-0139492.
XX			99US-0139454.
XX			99US-0139455.
XX			99US-0139456.
XX			99US-0139457.
XX			99US-0139458.
XX			99US-0139459.
XX			99US-0139460.
XX			99US-0139461.

PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

QY	29	IPVIETLIKEEIPFGLNITGYTLKFLPKDIIDLVKGGIASDL-----IEIIGTSYTHAIL	83
Dd	256	LPILE-MVPEELHF---TVPYALSFIMGDPDKMATLGIDNQLTGVKIEKLQRQLTKTML	311
QY	84	PLL-----PLSRVEAQVQRDRVKEELEFEVSPKGFWLPEL	118
Dd	312	PLLSELGIIPRETLWLKLLRSGCAYANSRIHA-VQ-----AEVLVLASGKDMMLPSQ	365
QY	119	AYDPIIPAIL-----KDNGYEYLFADGEAMLFSAHLSAIPKIPLYPHLIKAQREK	170
Dd	366	EAKRLHGLLKNCVRCFKDNGHTLLEDISILL-----TVIKGTG-----KYRRSW	412
QY	171	RFRYSIS-----YLLG--LRLEIRKAIKLVF-----EGKVTLKAVKDIEAVP---	209
Dd	413	RYDLVSDFLPSPSKGELAYALDEVYLGELRNNAVGSVFFFSTMEDGKI-----VKGLAGVDPKGP	468


```

QY      210 WVAVNTAVMLGIGRLPLMNPKKVASWIEDKDNILL-----YGTDIE 250
Db      469 VLLVGYYHLMGLMGLELGPMS-----AFIKEK-NILFRGMAHPVLVSDNDPAKAFDYGDWIK 522

QY      251 FIGYRDIAGYRMSVEGLLEVIDELNSELCLPS---ELKHSGRELYLRTSSWAPDKSLRIW 307
Db      523 VFG-----AYPVTATNLFKLLDSKSHVLLFPGGAREALHNRGEOY-----KLIIW 566

QY      308 REDEGNARL 316
Db      567 PEQQEFVRM 575
```

RESULT 12

AAG48653
ID AAG48653 standard; Protein; 726 AA.

AC AAG48653;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 61463.

Protein identification; signal transduction pathway; metabolic pathway;
KW
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
KW

OS *Arabidopsis thaliana*.

PN EP1033405-A2.

PD 06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	07-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-037528.

PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.

CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes, themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1073 AA;
SQ
Query Match 5.5%; Score 103; DB 22; Length 1073;
Best Local Similarity 21.2%; Pred. No. 0.77;
Matches 58; Conservative 48; Mismatches 90; Indels 78; Gaps 15;

QY 107 EVSPKGFWLPPELAYDPIIPAILKNDNGYEYLFADGEAMLFSAHLNSAIKPIKPLYPHLIKA 166
Db 309 EISP-----YYDSL---LVKLSTHAISFKQAEKVMVRSRLREMRIRGVKTNIPFLINV 357
QY 167 QREKRF-----RYISYLLGLRELKAI-----KLVFEGKVTCLKAVKDIE----- 205
Db 358 MNKKFTSGDYTTKFIETPELFDIQPSLDRGKTLEYIGNVTINGFPNVEKRPKPDYEL 417
QY 206 -AVPVVAVNTAVMLGIGR-LPLMNPKKVASWIEDKNILLYGTDIEF-IGYRDIAGYRM 262
Db 418 ASIPTVSSSKIASFSGKQLLDEVGPKGVAEWVKQDDVLL--TDTTFRDAHQSLLATRV 475
QY 263 SVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAPDK-SLRIWREDEGNARLNMLSY 321
Db 476 RTKDMINIAS-----KTADVFKDGFSLMW-----GGATFD-VAY 509
QY 322 NMRGELAFLAENSDARGWEPLPERRLDAFRAIYN 355
Db 510 N-----FLKENP-----WERLERLR-----KAIPN 529

RESULT 14
AAU33972
ID AAU33972 standard; Protein; 1147 AA.
XX
AC AAU33972;
XX
DT 14-FEB-2002 (first entry)
DE Staphylococcus aureus cellular proliferation protein #248.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR N-PSDB; AAS51831.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 5468; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes, themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1147 AA;
SQ
Query Match 5.5%; Score 103; DB 22; Length 1147;
Best Local Similarity 21.2%; Pred. No. 0.85;
Matches 58; Conservative 48; Mismatches 90; Indels 78; Gaps 15;

QY 107 EVSPKGFWLPPELAYDPIIPAILKNDNGYEYLFADGEAMLFSAHLNSAIKPIKPLYPHLIKA 166
Db 385 EISP-----YYDSL---LVKLSTHAISFKQAEKVMVRSRLREMRIRGVKTNIPFLINV 433
QY 167 QREKRF-----RYISYLLGLRELKAI-----KLVFEGKVTCLKAVKDIE----- 205
Db 434 MNKKFTSGDYTTKFIETPELFDIQPSLDRGKTLEYIGNVTINGFPNVEKRPKPDYEL 493
QY 206 -AVPVVAVNTAVMLGIGR-LPLMNPKKVASWIEDKNILLYGTDIEF-IGYRDIAGYRM 262
Db 494 ASIPTVSSSKIASFSGKQLLDEVGPKGVAEWVKQDDVLL--TDTTFRDAHQSLLATRV 551
QY 263 SVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAPDK-SLRIWREDEGNARLNMLSY 321
Db 552 RTKDMINIAS-----KTADVFKDGFSLMW-----GGATFD-VAY 585
QY 322 NMRGELAFLAENSDARGWEPLPERRLDAFRAIYN 355
Db 586 N-----FLKENP-----WERLERLR-----KAIPN 605

RESULT 15
AAG82528
ID AAG82528 standard; Protein; 360 AA.
XX
AC AAG82528;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2150.
XX

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 07:24:24 ; Search time 18 Seconds
(without alignments)
594,997 Million cell updates/sec

Title: US-09-619-032-4
Perfect score: 1880
Sequence: 1 MRALVFHGNLQYAEIPKSEI.....RRLDFAIRYNDWRGNGEP 364

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1587	84.4	346	US-08-613-220B-4	Sequence 4, Appli
2	140.5	7.5	647	US-07-894-212A-8	Sequence 8, Appli
3	140.5	7.5	649	US-07-894-212A-2	Sequence 2, Appli
4	140.5	7.5	650	US-07-893-928A-1	Sequence 1, Appli
5	101	5.4	329	US-08-270-013B-2	Sequence 2, Appli
6	101	5.4	329	US-08-838-418-2	Sequence 2, Appli
7	94.5	5.0	367	US-09-134-001C-3678	Sequence 3678, Ap
8	91.5	4.9	1009	US-09-693-146-4	Sequence 4, Appli
9	90.5	4.8	406	US-09-134-001C-3570	Sequence 3570, Ap
10	88.5	4.7	764	US-09-235-451-36	Sequence 36, Appl
11	88.5	4.7	2089	US-08-418-893D-23	Sequence 23, Appl
12	88.5	4.7	2089	US-08-418-893D-24	Sequence 24, Appl
13	88	4.7	1154	US-09-134-001C-3428	Sequence 3428, Ap
14	87.5	4.7	689	US-08-248-021A-2	Sequence 2, Appli
15	87	4.6	235	US-09-141-135-2	Sequence 2, Appli
16	87	4.6	657	US-09-355-166-1	Sequence 1, Appli
17	85.5	4.5	511	US-09-134-078-21	Sequence 21, Appl
18	84	4.5	636	US-09-564-805-237	Sequence 237, App
19	83.5	4.4	891	US-09-134-001C-4913	Sequence 4913, Ap
20	83	4.4	1110	US-08-118-441-29	Sequence 29, Appl
21	83	4.4	1110	US-08-338-579A-29	Sequence 29, Appl
22	83	4.4	1110	PCT-US94-09851-29	Sequence 29, Appl
23	82.5	4.4	549	US-08-676-279-59	Sequence 59, Appl
24	82.5	4.4	788	US-08-907-166-6	Sequence 6, Appli
25	82	4.4	303	US-09-002-298-1	Sequence 1, Appli
26	81.5	4.3	806	US-08-451-715A-6	Sequence 6, Appli
27	81	4.3	550	US-08-484-493-2	Sequence 2, Appli

28	81	4.3	550	1	US-08-484-494-2	Sequence 2, Appli
29	81	4.3	550	2	US-08-345-212-2	Sequence 2, Appli
30	81	4.3	550	4	US-09-249-003-2	Sequence 2, Appli
31	80.5	4.3	491	1	US-08-489-733-3	Sequence 3, Appli
32	80.5	4.3	491	2	US-08-993-581B-3	Sequence 3, Appli
33	80.5	4.3	641	1	US-07-718-535-3	Sequence 3, Appli
34	80.5	4.3	641	1	US-08-161-999-3	Sequence 3, Appli
35	80.5	4.3	1676	4	US-08-487-283A-2	Sequence 2, Appli
36	80	4.3	329	4	US-09-180-827-9	Sequence 9, Appli
37	80	4.3	506	2	US-08-849-480A-5	Sequence 5, Appli
38	80	4.3	541	4	US-09-627-376-11	Sequence 11, Appl
39	80	4.3	1065	4	US-09-412-545-2	Sequence 2, Appli
40	79.5	4.2	906	2	US-08-609-230A-9	Sequence 9, Appli
41	79.5	4.2	1751	4	US-09-136-574A-44	Sequence 44, Appl
42	79	4.2	246	4	US-09-134-001C-4606	Sequence 4606, Ap
43	79	4.2	623	4	US-09-104-068-4	Sequence 4, Appli
44	79	4.2	637	4	US-09-104-068-2	Sequence 2, Appli
45	79	4.2	889	4	US-09-132-316-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-613-220B-4
; Sequence 4, Application US/08613220B
; Patent No. 5958751
; GENERAL INFORMATION:
; APPLICANT: Murphy, Dennis
; APPLICANT: Reid, John
; TITLE OF INVENTION: ALPHA-GALACTOSIDASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,220B
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-678-5070
; TELEFAX: 619-68-5099
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 346 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-613-220B-4

Query Match 84.4%; Score 1587; DB 2; Length 346;
Best Local Similarity 94.8%; Pred. No. 2.3e-160;
Matches 345; Conservative 1; Mismatches 0; Indels 18; Gaps 18;
QY 1 MRALVFHGNLQYAEIPKSEIPKVIEKAYIPVETLIKEIPFGLNITGYTLKFLPKDIID 60

Db 1 LRALVFHGNLQYAEIPKSE-PKVIEWKAYIPVETLIKKEE-PFGLNITGYTLKFLPKDII- 57
QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVORDREVKEELFEVSPKGFWLPCLAY 120
Db 58 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVORDREVKEELFEVSPKGFWLPCLAY 114
QY 121 DPIPAILKDNNGEYELFADGEAMLFSAHLSAIPKIPKPLYPHLIKAKREKFRYISYLLG 180
Db 115 DPIPAILKDNNGEYELFADGEAMLFSAHLSAIPKIPKPLYPHLIKAKREKFRYISYLLG 171
QY 181 LRELKAKILVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
Db 172 LRELKAKILVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 228
QY 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db 229 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 285
QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLEAENSDARGWEPLPERRLDFAFRAIYNDWRGE 360
Db 286 DKSLRIWREDEGNARLNMLSYNMRGELAFLEAENSDARGWEPLPERRLDFAFRAIYNDWRG- 342
QY 361 NGEF 364
Db 343 NGEF 346

RESULT 2
US-07-894-212A-8
; Sequence 8, Application US/07894212A
; Patent No. 5366883
; GENERAL INFORMATION:
; APPLICANT: ASADA, KIYOZO
; APPLICANT: UEMORI, TAKASHI
; APPLICANT: MUKAI, HIROYUKI
; APPLICANT: KATO, IKUNOSHIN
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/894,212A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95469/C-1195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-894-212A-8

Query Match 7.5%; Score 140.5; DB 1; Length 647;
Best Local Similarity 22.8%; Pred. No. 4.4e-06;
Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;
QY 25 EKAYIPVETLIKKEIP---FGLNITGYTLKFLP---KDIIDLKVGKGIASDLIEIIGTSY 78
Db 27 EKCVWPFLETL---EYPNMKVAIHTSGPLIEWLQDNRPYIDLLRSVLKRGQVEIVVAGF 84
QY 79 THAILPLPLSRVEAQVORDREVKE---ELFEVSPKGFWLPCLAYDPIIPAILKDNNGEYEL 136
Db 85 YEPVLASIP---KEDRIEQIRLMKEWAKSIGDFDARGVWLTERTVWQPELVKTLKESGIDYV 141
QY 137 FADGEAMLFSAHLSAIPKIPKLY-PHLIKAKRE-----KFRYISYLLGLRELKRA 187
Db 142 IVD-----DYHMSAGLSKEELYWPYYTDEGGEVIAVFPIDEKLR---YLIPFRPVDRK 192
QY 188 IKL---VFEGKVTLKAV---KDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWI----- 236
Db 193 LEYLSLIDGDESKVAVFHDGKFGIWPPTYENVY-----EKGLWREFFD 239
QY 237 ---EDKDNILLYGTDIE-----FIGYRDIAGYRMSVEGLLEVIDELNSELCLP----- 281
Db 240 ISSDEKINMLYTEILEYKPRGLVYLPYASV-----FEM-----SEWSLPKQARLF 287
QY 282 ---SELKHSG---RELRLTSSWAPDKSLRIWREDEGN---ARLNMLSYNMRGELAF 330
Db 288 VEFVNELKVGKIFEKYRVRGGIW---KNF-FYKYPESNYMHKRLMVLKLVN--- 339
QY 331 AENSDARGWEPLPERRLDFAFRAIYND---WRG 359
Db 340 -----PEARKYLLRAQCNDAYWHG 358

RESULT 3
US-07-894-212A-2
; Sequence 2, Application US/07894212A
; Patent No. 5366883
; GENERAL INFORMATION:
; APPLICANT: ASADA, KIYOZO
; APPLICANT: UEMORI, TAKASHI
; APPLICANT: MUKAI, HIROYUKI
; APPLICANT: KATO, IKUNOSHIN
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/894,212A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95469/C-1195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 649 amino acids
; TYPE: amino acid

RESULT 8
US-09-693-146-4
; Sequence 4, Application US/09693146
; Patent NO. 6413758
; GENERAL INFORMATION:
; APPLICANT: Xu, Shuang-yong

Db 556 ---LSQEARPEAPTGNATESVQPMEGQDEGNGAQYRGILEASLELPKFTTIGMELAF 612

RESULT 11

US-08-418-893D-23

; Sequence 23, Application US/08418893D

; Patent No. 5559220

; GENERAL INFORMATION:

; APPLICANT: ROESSLER, PAUL G

; APPLICANT: OHLROGGE, JOHN B

; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A

; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTOLLA CRYPTICA

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY

; STREET: 1617 Cole Blvd.

; CITY: Golden

; STATE: CO

; COUNTRY: USA

; ZIP: 80401-3393

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/418,893D

; FILING DATE: April 7, 1995

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/104,938

; FILING DATE: September 14, 1993

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: O'CONNOR, EDNA

; REGISTRATION NUMBER: 29,252

; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 303-231-1000

; TELEFAX: 303-231-1098

; TELEX:

; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2089 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-418-893D-23

Query Match 4.7%; Score 88.5; DB 1; Length 2089;

Best Local Similarity 19.6%; Pred. No. 9.9;

Matches 70; Conservative 65; Mismatches 110; Indels 113; Gaps 21;

QY 9 NLOYAEIPKSEIPKVIKAYIPVETLIK--EEIP-----FGLNITGYTLKFLPKDIIDL 61

Db 1057 SFQFADVPAKD--RVTRQGFVSIDDASKFAQQLPEILNSFGSKIAGDASKEGPNVNLQV 1114

QY 62 VKGGIASDL-IEII--GTSYTHAILPLPLSRVEAQVQRDREVKEELFEVSPKGFWLPEL 118

Db 1115 --GALSGDISIEDLEKATSANKDKLNLGVRTVTALIPRGKK-----DPSYYSFPQC 1164

QY 119 A---YDPIIPAILKNDNGYEYLFADGEAMLFSAHLSAIPKIPLYPHLIKAQR-EKRF-- 172

Db 1165 SGFKEDPL-----RRGMRPTFHHLLLELGRLENFAL 1195

QY 173 -----RYSYLLGLREL--RKAIKLVFEGKVTLLKAVKDIEAVPVVAVNTAVMLGIGR 223

Db 1196 ERIPAVGRNVQIYVGSEKTAARRNAQVVF-----LRAISHTPGLTTFSGARRALLQGLDE 1250

QY 224 LPLMNPKKVASWIEDKDNILLYG-----TDI-----EFIGYRDIAGYRMSVEGLLEVID 272

Db 1251 LERAQANSKVS-VQSSSRYYLHSLPEQSDATPEEIAKEFEG-----VID 1293

QY 273 ELNSELCLPSELKHSGRYLYRTSSWAPDKSLRIWREDE-GNAR---LNMLSYNNMRGE 326

Db 1294 KLKSRL-----AQLTKLRVDE--IETKVRVTQDEDEGSPRVVPVRLVASSMQGE 1341

RESULT 12

US-08-418-893D-24

; Sequence 24, Application US/08418893D

; Patent No. 5559220

; GENERAL INFORMATION:

; APPLICANT: ROESSLER, PAUL G

; APPLICANT: OHLROGGE, JOHN B

; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A

; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTOLLA CRYPTICA

; NUMBER OF SEQUENCES: 25

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY

; STREET: 1617 Cole Blvd.

; CITY: Golden

; STATE: CO

; COUNTRY: USA

; ZIP: 80401-3393

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/418,893D

; FILING DATE: April 7, 1995

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/104,938

; FILING DATE: September 14, 1993

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: O'CONNOR, EDNA

; REGISTRATION NUMBER: 29,252

; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 303-231-1000

; TELEFAX: 303-231-1098

; TELEX:

; INFORMATION FOR SEQ ID NO: 24:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2089 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: N-terminal

US-08-418-893D-24

Query Match 4.7%; Score 88.5; DB 1; Length 2089;

Best Local Similarity 19.6%; Pred. No. 9.9;

Matches 70; Conservative 65; Mismatches 110; Indels 113; Gaps 21;

QY 9 NLOYAEIPKSEIPKVIKAYIPVETLIK--EEIP-----FGLNITGYTLKFLPKDIIDL 61

Db 1057 SFQFADVPAKD--RVTRQGFVSIDDASKFAQQLPEILNSFGSKIAGDASKEGPNVNLQV 1114

QY 62 VKGGIASDL-IEII--GTSYTHAILPLPLSRVEAQVQRDREVKEELFEVSPKGFWLPEL 118

Db 1115 --GALSGDISIEDLEKATSANKDKLNLGVRTVTALIPRGKK-----DPSYYSFPQC 1164

QY 119 A---YDPIIPAILKNDNGYEYLFADGEAMLFSAHLSAIPKIPLYPHLIKAQR-EKRF-- 172

Db 1165 SGFKEDPL-----RRGMRPTFHHLLLELGRLENFAL 1195


```

Query Match      4.6%; Score 87; DB 2; Length 235;
Best Local Similarity 21.0%; Pred. NO. 0.42;
Matches 49; Conservative 37; Mismatches 71; Indels 76; Gaps 11;

QY 165 KAQREKRRFYISYLLGLRELRLKAIKLVFEGKVTLKAVKDIEAVPVWVAVNTAVMLGIGRL 224
      | : : : | | : : : | : : | : : | | : : | | : : |
Db 27 KSNNQKRFN-----EEQIKSLELIFESETRLEPRKKVQ-----VARELG----- 65

QY 225 PLMNPKKVASWIEDKD-----NILLYGTDIEFIGYRDIAG-----YRMSVEGLL 268
      : : : : | : : : | | | | | : : : | : : : | : : |
Db 66 --LQPRQMTIWFQNKRAKRWKTKQLEKEYNTLR-----ANYNNIASQFEIMKKEKQSLV 116

QY 269 EVIDELNSELCLPSELKH-----SGRELYLRTSSWAPD-KSLRIWREDEGNARLNLSYN- 322
      : : : : | | : : | : : : | : : : | : : : | : : |
Db 117 SELQRLNEEMQRPKEEKHHECCGDQGLALSSSTESHNGKSEPEGRLDQGSVLCNDGDYNN 176

QY 323 -----MRG-----ELAFLAENSDARGWEPLPERRLDAFRAIYNDRWGENGE 363
      : : | | : : | : | : | : | : | : | : | : |
Db 177 NIKTEYFRVQGETDHELMNIVEKAD-----DSCLTSSENWGGFNDS 217

```

Search completed: November 18, 2002, 07:26:22
Job time : 23 secs

QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPY 120
Db 61 LIKEGIESGLIEILGTSYTHAILPLPLSRVEAQIKRDREVKENILEVSPKGFWLPY 120
QY 121 DPIIPAILKDNQYEVYLFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRYISYLLG 180
Db 121 DPIIPAILRDNVYEVYLFADGEAMLFNSHLNSAIKPIKPLYPHLIKAQREGGLVYLLG 180
QY 181 LRELKKAIKLVFEGKVTILKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
Db 181 LRELKKAINKLVFEGKVTLEAVKEIEAIPVWVSINTAVMLGAGRPPLMNPKKVAKWVKEKD 240
QY 241 NILLYGTDIEFTGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db 241 EILLYGTDIEFLGYRDIAGYKITISNLEIINELEGELGLPRKIKHSEKKLYLRTSSWAP 300
QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSADARGWEPLPERRLDFAIRYNDWRGE 360
Db 301 DKSLRIWTEDEGNARLNMLTSYMDGELAFLAENSADARGWEPLPERRLDFAKAIYTHRSE 360
QY 361 NGE 363
Db 361 NGK 363

RESULT 2

ID O58106 PRELIMINARY; PRT; 364 AA.
AC O58106;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein PH0368.
GN PH0368.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kwarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000002; BAA29442.1; -;
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 364 AA; 41755 MW; 7B4B36AB4A975BAD CRC64;

Query Match 80.1%; Score 1506; DB 17; Length 364;
Best Local Similarity 76.6%; Pred. No. 5.8e-112;
Matches 278; Conservative 47; Mismatches 38; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKEIPFGLNITGYTLKFLPKDIID 60
Db 1 MRALIFHGNLQYAEIPKHEISKVIEKSYFPTISELIKREIPFGLNITGYSLOFLPQELIH 60
QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPY 120
Db 61 LIKEGESELIEILGTSYTHAILPLPLTLRIEAQIKRDREIKEIFEVSPKGFWLPY 120
QY 121 DPIIPAILKDNQYEVYLFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRYISYLLG 180
Db 121 DPIIPAILRDNVYEVYLFADGEAMLFNSHLNSAIKSIKPLYPYLIKAQREGGCVYLLG 180

QY 181 LRELKKAIKLVFEGKVTILKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
Db 181 LRELKKAINKLVFEGKVTLEAVKDIEAIPVWVSINTAVMLGAGRPPLMSPKRVANWIKGD 240
QY 241 NILLYGTDIEFTGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db 241 EILLYGTDIEFLGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSADARGWEPLPERRLDFAIRYNDWRGE 360
Db 301 DKSLRIWKEDEGNARLNMLTYCMDGEFAFLAENSADARGWEPLPERRLDFAKAIYKYWRNE 360
QY 361 NGE 363
Db 361 NGK 363
RESULT 3
Q8YUA2
ID Q8YUA2 PRELIMINARY; PRT; 529 AA.
AC Q8YUA2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein Alr2450.
GN ALR2450.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003589; BAB74149.1; -;
DR InterPro; IPR003803; DUF200.
DR Pfam; PF02651; DUF200; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 529 AA; 62103 MW; 6AB10681738C8805 CRC64;
Query Match 8.9%; Score 167; DB 16; Length 529;
Best Local Similarity 21.0%; Pred. No. 6.3e-05;
Matches 104; Conservative 55; Mismatches 162; Indels 174; Gaps 21;
QY 3 ALVFHGNLQYAEIPKSE-----IPKVIKAYIP---VIETLIKEIPFGL----- 44
Db 7 ALVLHAHLPEFVRHPESDYVLEEEWLYEAITETIPLLLKVFGLKRDGIDFKITMSMTPL 66
QY 45 -----NITGYTLKFLPK----- 56
Db 67 VSMLRDPPLQERYDAHLAQLEELIELEGERNAQNGHLRYLAEHVATEFNEARQMWERYNG 126
QY 57 DIIDLKVGGLASDLIEIIGTSYTHAILPLPL--SRVEAQVQDRREVKEELFEVSPKGF 114
Db 127 DLVTAFKQFQDSNNLEIITCGATHGYLPLMKMYPQAVWAQIQVACEHYEETFGRRPKGIW 186
QY 115 LPelayDPIIPAILKDNQYEVYLFADGEAMLFSAHLNSAIKPIKPL----- 159
Db 187 LPECAVYEGLERMLADAGLRYFLTDGHLILY-ARPRPRFGTYAPIFTETGVAAGRDHES 245
QY 160 -----YPHLIKAQR-----EKFRFYISYLLGLRELKKAIKLVFEGKVT----- 197
Db 246 SQQVWSSEVGYPGAAEYREFYKDLGWEAEYIKPYIMPNGORKNTGIKYH-KITGRGLG 304
QY 198 --LKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKDNL--LYGT----- 247
Db 305 LSDKALYD----PYWAKEKA-----EHAANFMYNRERQAEHLYGIMQRPPI 347

QY 248 -----DIEFIGYRDIA-----YRMSVEGLLEVIDELNSE-----LCLPSELKH 286
Db 348 IVSPYDAELFGHWYEGPWFIIDYLFKRSWYDOGTAMTHLADYLRNEPTQVCRPSQ--- 404
QY 287 SGRELYLRTSSWAPDKSLRIWREDEGNA-----RLNMLSYNMRGELAFLAENS DARGWEPL 342
Db 405 -----SSWGKGFHEYWL-NETNANTYPHLHKAARMI-EISTL-EPEDELGWRAL 452
QY 343 PERRLDAFRAIYNDW 357
Db 453 NQAARELLLAQSSDW 467
RESULT 4
P74630
ID P74630 PRELIMINARY; PRT; 529 AA.
AC P74630;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein sll0735.
GN SLL0735.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90916; BAA18743.1; -;
DR InterPro; IPR003803; DUF200.
DR Pfam; PF02651; DUF200; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 529 AA; 62075 MW; 4C1A45048A784E30 CRC64;
Query Match 8.6%; Score 161.5; DB 16; Length 529;
Best Local Similarity 21.4%; Pred. No. 0.00017;
Matches 101; Conservative 47; Mismatches 140; Indels 183; Gaps 19;
QY 3 ALVPHGNLOYAEIPKSE-----IPKVEIKAYIPVI---ETLIKEIPFGLNIT----- 47
Db 7 ALVLHAHLPFVRHPESDYVLEEWLYEAITETIPIIHVFEGKLRDGVDFKITSMTPL 66
QY 48 -----GYTLKFLPK----- 56
Db 67 VSMRLRDPQLQRYEAHLSLLOELLAKEIVRNEHNGHLOYLADFYAKEFAAIRETWERYDG 126
QY 57 DIIDLKGGIASDLIEIIGTSYTHAILPLPL--SRVEAQVQRDREVKEELFEVSPKGF 114
Db 127 DLVTAFKQFQDSNNLEIITCGATHGYFPLMKYPQAVWAQIKVACEHYEENFGSPKGIW 186
QY 115 LPELAYDPIIPAILKONGEYLFADGEAMLFSA-----HLNS 151
Db 187 LPECAVEGVERMLADAGLRYFLVDGHGILYARPRPKFGTYAPIFTETGVAAFGRDHESS 246
QY 152 -----AIKPIKPLPHLIK-AQREKFRFYISYLLGLRELKAIKLVF-----EGKVT 197
Db 247 QQVWSSQVGYPGDPVYREFYKDLGWAEYKPYIMPNGQKNIGIKYHKITSRDGGLS 306
QY 198 LKAVKDIEAVPVVAVNTAV-----MLGI-GRPLPLNPKKVAS-----WI 236
Db 307 EKAWDY----PYWAKEKAAEHASNFMYNRQQVGHLSGIMGRPPLVSPYDAELFGHWY 362
QY 237 E-----DKDNI-----LLYGTDIEFI-----GYRDIAGY----- 260

Db 363 EGPWFIDYLFKRSWFDQDTFEMTHLADYLRGNPHQVCRPSQSSWGKGFHEYWLNDTNA 422
QY 261 -----RMSVEGLLEVIDELNSELCLPSELKHSGRELYL-RTSSWA 299
Db 423 WIYPHLHKAARMIELSHREAVDELEEK-----ALNQAARELLLAQSSDW 468
RESULT 5
Q97BM4
ID Q97BM4 PRELIMINARY; PRT; 378 AA.
AC Q97BM4;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Alpha-amylase.
GN TV0431 OR TVG0421416.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GSSI / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Ohta Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaeal adaptation to higher temperatures revealed by genomic
sequence of Thermoplasma volcanium.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL; AP000992; BAB59573.1; -;
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57; 1.
KW Complete proteome.
SQ SEQUENCE 378 AA; 44957 MW; 49FCCF63BEE6B3A7 CRC64;
Query Match 8.1%; Score 152.5; DB 17; Length 378;
Best Local Similarity 21.4%; Pred. No. 0.00056;
Matches 67; Conservative 48; Mismatches 123; Indels 75; Gaps 11;
QY 22 KVIEKAYIPVIETLIKEIPFGLNITG-----YTLKFLPKDIIDLKGGIASDLIEI 73
Db 42 RVAERSYIPATRNLMYGIKSSFSITGTAVEQALMYTK-----VIDAIDDYVKSGLCM 96
QY 74 IGTSYTHAILPLPLSRVEAQVQRDREVKEELFEVSPKGFWLPYLDPIIPAILKDN 133
Db 97 LSETYYHSLASIWNYDEFKQVDMHRLMKRIFNVVVKVFRNTELYDDRIAEMVKRMGF 156
QY 134 EYLFADGEAMLFSAHLNSAIKPIKPLPHLIKAREKFRFYIS----YLLGLRELKAIKL 190
Db 157 TSIITEG-----TDSIVKDHSPNY-----RYASPSGLNLYLRNYVMSDNI 196
QY 191 VFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKDNILLYGT 250
Db 197 SFRFSNT--KWKDYP-----LTADKYAKWINESEGDVNV---L 229
QY 251 FIGYRDIAGYRMSVEGLLEVIDEL-----NSELCLPSELKHSGRELYL--RTSSWA 299
Db 230 FMDYETTFGEHQETGTGIFDFMKYLPVYFRDYGIETITISEAEKRRHVYKDVLSIPETISWA 289
QY 300 PD-KSLRIWREDE 311
Db 290 DTRDLSAWLENE 302
RESULT 6
Q97GF3
ID Q97GF3 PRELIMINARY; PRT; 527 AA.
AC Q97GF3;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE	Uncharacterized conserved protein, related to
DE	alpha-amylase/alpha-mannosidase.
GN	CAC2414.
OS	Clostridium acetobutylicum.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC	Clostridiales; Clostridiaceae; Clostridium.
OX	NCBI_TaxID=1488;
RN	{1}
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX	MEDLINE=21359325; PubMed=11466286;
RA	Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA	Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA	Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA	Bennett G.N., Koonin E.V., Smith D.R.;
RT	"Genome sequence and comparative analysis of the solvent-producing
RT	bacterium Clostridium acetobutylicum.";
RL	J. Bacteriol. 183:4823-4838(2001).
DR	EMBL; AE007742; AAK80369.1; -
DR	InterPro; IPR003803; DUF200.
DR	Pfam; PF02651; DUF200; 1.
KW	Complete proteome.
SQ	SEQUENCE 527 AA; 61485 MW; 03706E16907929C6 CRC64;

Query Match	8.0%;	Score 151;	DB 16;	Length 527;
Best Local Similarity	18.9%;	Pred. No. 0.0012;		
Matches 92;	Conservative 63;	Mismatches 119;	Indels 212;	Gaps 20;

QY	5	VFHGNLOYAEIPKSE-----IPKVIEWKAYIPVIE--TLIKEEIPFG-----	43
		: : : : : : : : : : : : : : : :	
Db	9	VLHSHMPFVRHPETRDSLEERWLFTEAMSECYIPLIEVDNLLKDNIKFRMTMSITPPLMS	68
QY	44	-----LNITGYTLKFLPKDII-----DL 61	
		: : : : : : : : : : : : : : :	
Db	69	MLQDEYLSRYLNYLKKTIELSEKEILRTKNNREENKVALFYNKRAENTLKIYEKYDNNL	128
QY	62	VKGGIASD---LIEIIGTSYTHAILPLLPLSR--VEAQVQRDREVKEELFEVSPKGFWLP	116
		: : : : : : : : : : : : : : : : : : : : :	
Db	129	INAFKRYDRLGCVIEITCAATHALLPLILINQAVKACIATGVQSYINTMGHEPNGIWL	188
QY	117	ELAYDPIIPAILKNDNGVEYLFADGEAMLFSAHLNSAIKPKIPLYPHLIKAQREKFRYIS	176
		: : : : : : : : : : : : : : : : : :	
Db	189	ECAYTYGIDNILSEFGIKYFISEGKAIDYASP-----KPMY-----	224
QY	177	YLLGLRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWI	236
		: : : : : : : : : : : : : : : : : : : :	
Db	225	-----GTNT-----PIAAPSGVCAFG 240	
QY	237	EDKDNILLYGTDIEFIG-----YRDIAGYRMSVEGLLEVIDELNSELCLPSELKH	286
		: : : : : : : : : : : : : : : : : : : :	
Db	241	RDMDS--SYQVWSDFMGYPGDENYREFYRDI-GFELPMYIKPYINE--NGIRIDTGFKY	295
QY	287	-----SGRE-LYLRTSS---WAPDKSLRIWREDEGN-ARLNM-----	318
		: : : : : : : : : : : : : : : : : : : :	
Db	296	YKITNSGEGKGYNRENAMKKVWEHASHFASCRHDQINAAAAANMDKPPITCPTYDTLYG	355
QY	319	-----LSYNMRGELAFLAENSADARGWEPLPERRLDAFRATYNDW 357	
		: : : : : : : : : : : : : : : : : : : :	
Db	356	HWFEFGPDFINAFIRKSAEDWTSYELITPTTEYLKNNSMVQCSSP-----SSW	404
QY	358	RGENGE 363	
		:	
Db	405	-GENGD 409	

RESULT 7	
Q973T0	PRELIMINARY; PRT; 443 AA.
ID Q973T0	
AC Q973T0;	
DT 01-DEC-2001 (TrEMBLrel. 19, Created)	
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)	
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
DE Hypothetical proteinST0817.	

[illegible]

RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain 7.";
RL DNA Res. 8:123-140(2001).
DR EMBL: AP000985; BAB66135.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 895 AA; 101064 MW; 6BC7CD83D0DDFBC CRC64;

Query Match 7.5%; Score 140.5; DB 17; Length 895;
Best Local Similarity 21.0%; Pred. No. 0.017;
Matches 85; Conservative 72; Mismatches 163; Indels 85; Gaps 18;

QY 3 ALVFHGNLQYAEIPKSEIPKVIKAYIPVI---ETLIKEE-----IPFGLNIT----- 47
Db 377 SLVGAYELQALLIKQFNVSVDID--FTPVLVYQWETILHEENATFTSNFGVNVSHDIEAV 434
QY 48 GYTLKFLPKDIIDLKGGIASDLIEIIGTSYTHAILPLL---PLSRVEAQVORDREVKE 103
Db 435 NYTL-----NLKTLIKEGKVEVLTVPFHPLOPLLLQDGYNSDVLQAIEMGENMTH 486
QY 104 ELFEVSPKGFWLPELAYDPIIPAILKNDNGYELFADGEAML-FSAHLNSAIKPIKPLYPH 162
Db 487 EVFGVWANGTWTPEMAFMDLVGLYNESGISFTILDQQAFPLPYVTLVNGSLNPDQPFIVE 546
QY 163 LIKAQREKFRYISVLLGLRELKAIKLVFECKVTLK---AVKDIEAVPVVAVNTAVML 219
Db 547 -----NNLQGTIIIVLFR-NTLSNEFGKFFSQSPQLTA--QELIQ 584
QY 220 GIGRLPLMNPKKVASWIEDKDNILLYG-----TDIEFIGYRDIAGYRMSVEGLLEVIDEL 274
Db 585 QLAEIYMNPGVVTVALDGENPLIFNPTTGPSDLYAI-YQALSEYQGS-----WLVTQT 638
QY 275 NSELCLPSELKHSRELYLRTSSWAPDKSLRIWR-----EDEGNARLNMLSYNMR 324
Db 639 ASEAI--ATHKPTSITNLVPNSW--DLNLNYYWNGYIGKTEIQWNSLAREYLIAYTVA 694
QY 325 -----GELAFLAENSARGWEPLPERRLDAFRALYNDWRGENGEP 364
Db 695 LGDNISPLVLPFNTPTNSTDLIDTLNLYLYVAEGSDWTWQTGPP 739

RESULT 9
Q8U136
ID Q8U136 PRELIMINARY; PRT; 632 AA.
AC Q8U136;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical protein Pf1393.
GN PF1393.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AE010242; AAL81517.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 632 AA; 74491 MW; C5A165981CB7A971 CRC64;

Query Match 7.2%; Score 135; DB 17; Length 632;
Best Local Similarity 21.9%; Pred. No. 0.029;
Matches 95; Conservative 52; Mismatches 148; Indels 138; Gaps 20;

QY 14 EIPKSEIPKVIKAYIPVIETLIKEIPFGLNITGYTLKFLPK----- 56
Db 71 EYMKKEFEKYITRKL-----ELMKED-----LTKYTNDQLKKAIAIMITYFEDVYRWK 119
QY 57 -----DIIDLKVGKGIASDLIEIIGTSYTHAILPLLPLSR-VEAQVORDREVKEELFEVSPK 111

Db 120 EINGDIIGKREFQEGAGYLEITSAATHGYLLPLGRDEAIEGQIANAIAKTYEYFQRRPR 179
QY 112 GFWLPELAYDPI-----IPAILKNDNGYELFADGEAMLFSAHLSAIAKP 155
Db 180 GMWLPELAYDPI-----IPAILKNDNGYELFADGEAMLFSAHLSAIAKP 238
QY 156 IKP-----LYPHLIK-----AQREKFRYISYLLG-----LRELKAIKLVFEKG 195
Db 239 ILPEKTKKSTLRPYFLKNGIAVFARNRETGIQVWSADIGYPCDFWYREFHK----- 289
QY 196 VTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCK-----VASWIED-- 238
Db 290 ---KAEK--SGQYWRVTSKEIDLG-GKEPYV-PEKAMERVEEHARHFISSLVSEEFK 342
QY 239 ----KDNILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSEL--KHSGR--E 290
Db 343 KKEGEYGIWVAPYDTFLFGHWFEQVKKW-LGKVLELAELKIGIKTTTISEFLDNFDGKRYE 401
QY 291 LYLRSSW-----APDKSLRIWREDEGNARLNMLSYNMRGEL 327
Db 402 IELPEGSWMGFGTHYTWNPVEVETWPIIHLAEDRMVALASKYLGRDELTDRLDQLGRE 461
QY 328 AFLAENSARGWE 340
Db 462 LLLIESSD---WQ 471

RESULT 10
O50094
ID O50094 PRELIMINARY; PRT; 560 AA.
AC O50094;
DT 01-JUN-1998 (TremBLrel. 06, Created)
DT 01-JUN-1998 (TremBLrel. 06, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Hypothetical protein PH1386.
GN PH1386.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kwarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL: AP000006; BAA30492.1; -.
DR InterPro; IPR003803; DUF200.
DR Pfam; PF02651; DUF200; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 560 AA; 65769 MW; 3F4CEF5267D167F8 CRC64;

Query Match 6.9%; Score 129.5; DB 17; Length 560;
Best Local Similarity 25.6%; Pred. No. 0.067;
Matches 60; Conservative 28; Mismatches 81; Indels 65; Gaps 12;

QY 14 EIPKSEIPKVIKAYIPVIETL-----IKEIPFGLNITGYTLKFLPKDIIDL---VK 63
Db 71 EYIKREFEKYMERKLSMEEDLERFKDEKLREAINF---MIGVF-----KDVYSYWKSID 122
QY 64 GGIASDL-----IEIIGTSYTHAILPLLPLSR-VEAQVORDREVKEELFEVSPKGF 114
Db 123 GNILGKFRELQDEGYVEVITSAATHGYLLPLGRDEAIEAQLNGIKYVEKYFGRKPRGIW 182
QY 115 LPELAYDPI-----IPAILKNDNGYELFADGEAMLFSAHLSAIAKPIKP 158

Db 183 LPECAAYRPGDLWKSPSTGEVKKWRKGIEHFLKFGIEYFFVE-----SHLIDK-GPVSL 234
QY 159 LYPHLIKAQREKFRFRYSYLLGLRELKRAIKLVFEGKVTLKAVKDIEAVPVVWA 212
Db 235 RYGNILPA-KTKRSTLRPYF-----LKNGIYAVFARNRET-----GIQVWSA 274

RESULT 11

Q9KD04 ID Q9KD04 PRELIMINARY; PRT; 923 AA.
AC Q9KD04;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein BH1415.
GN BH1415.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001512; BAB05134.1; -.
DR InterPro; IPR003803; DUF200.
DR InterPro; IPR001296; Glycos_transf_1.
DR Pfam; PF02651; DUF200; 1.
DR Pfam; PF00534; Glycos_transf_1;
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 923 AA; 106638 MW; 8C7B7D64F3603553 CRC64;

Query Match 6.8%; Score 127; DB 16; Length 923;
Best Local Similarity 23.8%; Pred. No. 0.22;
Matches 54; Conservative 31; Mismatches 70; Indels 72; Gaps 11;

QY 3 ALVFHGNLQYAEIPKSE-----IPKVEIKAYIPVIETLIKEIIPFGLNI----- 46
Db 7 SLVLHAHLPYVRHQEDRLERLWLFAMSETIYIPLLWALEKLPVKHAVTISFTPPVMEML 66
QY 47 -----TG-----TLKFLPKD-----IIDLVKG-----G 65
Db 67 SDPLVQTRYLNHLENTEQLLKKEKRTNDQRTQNLVQFYKQRYEKLKATFLOQDRNLLIG 126
QY 66 IASDLIE-----IIGTSYTHAILPLPLSLR-VEAQVQDRDREKVEELFEVSPKGFWLPELA 119
Db 127 FRS-LMENEQCILMTSAATHAFFPYLTKTEAIRAQVRHGIACFEQHFQKPLGLFWLPECA 185
QY 120 YDPIIPAILKDNQYELFADGEAMLF---SAHLNSAIKPIKPLY-PH 162
Db 186 FSPGVDRILFEEGIRYTFVDEHAVLTADPTPHKGSS-----APIYSPH 228

RESULT 12

Q9V294 ID Q9V294 PRELIMINARY; PRT; 1362 AA.
AC Q9V294;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE AMYLOPULLULANASE.
GN APU OR PAB0122.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ248283; CAB49104.1; -.
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57; 2.
KW Complete proteome.
SQ SEQUENCE 1362 AA; 154578 MW; 277AFAB4E14860D1 CRC64;

Query Match 6.7%; Score 126; DB 17; Length 1362;
Best Local Similarity 21.8%; Pred. No. 0.46;
Matches 83; Conservative 56; Mismatches 142; Indels 100; Gaps 16;

QY 32 IETLIKEIIPFGLNITGYTLKFLPKDIIIDLKGGIASDLIEIIGTSYTHAILPLLP---- 87
Db 236 VETVLKHOM-WLLNHT-----FEEHEKINLLG---NGNVEVTVPYTHPIGPILNDFGW 286
QY 88 LSRVEAQVQDRDREKVEELF---EVSPKGFWLPELAYDPIIPAILKDNQYELFADG---E 141
Db 287 YEDFDAQVKKANELYKEYLGAGKVTPKGGWAAESALNDKLTLEILAENGWKWYMTDQLVLE 346
QY 142 AMLFSAHLNSAIKP-----IKPLYPHLIKA 166
Db 347 KLGVPKTIESYYKPPWVAQFGDKKIYLFPRNHLDSDRVGFRYAGMNQYDAVKNFVEELKI 406
QY 167 QREKFRFYISYLLGLRELKRAIKLVFEGKVTLKAV-KDIEAVPVVAVNTAV-----ML 219
Db 407 OKQNYDGLSVYVITLDGENPWEHYPPFDGKLFLEELYRQLEELQKGLIRIVTPSEYIEMF 466
QY 220 G--IGRLPLMNPKKVASWIEDKDNILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSE 277
Db 467 GDKANKLTPKMMKRLDFTEDNVNALLKAKTL-----GELYDMVGVT----- 509
QY 278 LCLPSELKHSGRELYLRTSSWAPDKSLRIWREDEGNARLNMLSYNMRGELAFLAENSD-A 336
Db 510 -----EMQWPESSWI-DGTLSTW---IGEPQENIAWYWLALARKALFENKDNV 553
QY 337 RGWEPLPERRLDAFRAIYNDW 357
Db 554 KDNKAYEY---LFRAEGSDW 571

RESULT 13

Q8ZXX1 ID Q8ZXX1 PRELIMINARY; PRT; 471 AA.
AC Q8ZXX1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein PAE1048.
GN PAE1048.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
aerophilum.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009801; AAL63225.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 471 AA; 53638 MW; B7B379D3E214FB04 CRC64;

Query Match 6.6%; Score 124; DB 17; Length 471;

[illegible]

```

RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum.";
RL Nature 407:508-513(2000).
DR EMBL; AL445064; CAC11483.1; -.
KW Complete proteome.
SQ SEQUENCE 357 AA; 41845 MW; B8C1A97F4EB237F4 CRC64;

Query Match          6.5%; Score 122.5; DB 17; Length 357;
Best Local Similarity 16.6%; Pred. No. 0.13;
Matches 51; Conservative 59; Mismatches 126; Indels 71; Gaps

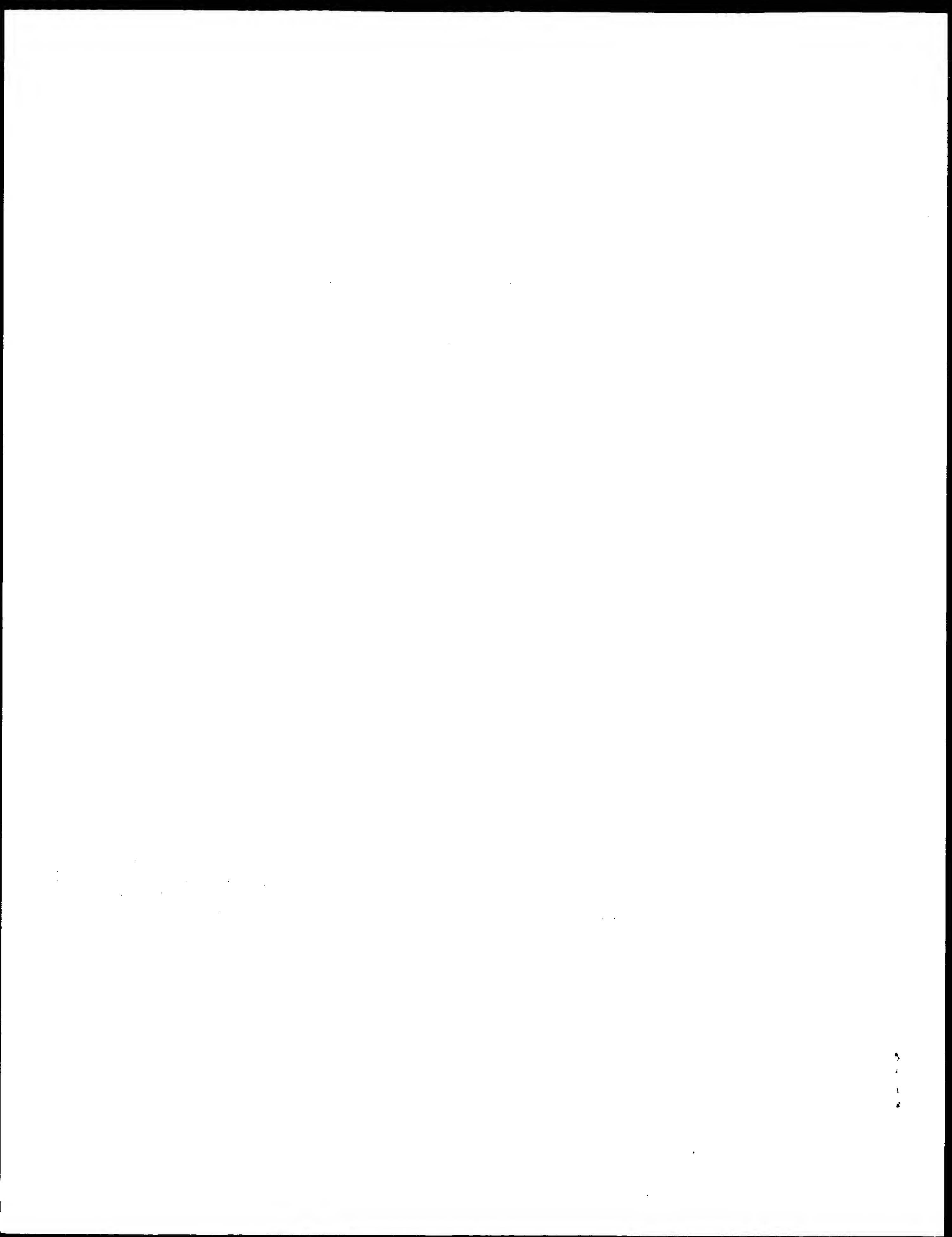
```

db 179 SNRAWADYPLTADKFARWISASSGEIVNL-----FM 209

QY 253 GYRDIAGYRMSVEGLLEVI-----DEINSELCLPSELKHSGREL--YLRTSSWA-P 300

Db	210	DYETFGHQRPETGTFELRYLPMYFEENDVHTILVREAEARHRTDRDFISVSKTTSWADK	265
QY	301	DKSLRIW	307
		::	
Db	270	NRDLSAW	276

Search completed: November 18, 2002, 07:27:04
Job time : 39 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 07:24:19 ; Search time 14 Seconds
(without alignments)
1078.385 Million cell updates/sec

Title: US-09-619-032-4
Perfect score: 1880
Sequence: 1 MRALVFHGNLOYAEIPKSEI.....RRLDAFRAIYNDWRGENGEP 364

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	161.5	8.6	633	1 AMYA_PYRHO	O57932 pyrococcus
2	159.5	8.5	685	1 AMYL_DICTH	P09961 dictyoglomu
3	146.5	7.8	467	1 AMYA_METJA	Q59006 methanococc
4	141.5	7.5	655	1 AMYA_PYRAB	Q9v298 pyrococcus
5	140.5	7.5	648	1 AMYA_PYRFU	P49067 pyrococcus
6	122.5	6.5	653	1 MALQ_PYRKO	O32450 pyrococcus
7	121.5	6.5	659	1 MALQ_THELI	O32462 thermococcu
8	109	5.8	311	1 MDH_BACSU	P49814 bacillus su
9	100.5	5.3	314	1 MDH_BACHD	Q9k849 bacillus ha
10	99	5.3	474	1 NOEK_RHISN	P55356 rhizobium s
11	98.5	5.2	705	1 RNR_AQUAE	O67834 aquifex ao
12	98.5	5.2	813	1 LON2_BORBU	O51558 borrelia bu
13	97.5	5.2	312	1 MDH_BACTC	Q9x4k8 bacillus th
14	96.5	5.1	471	1 UDPG_PYRPF	O64459 pyrus pyrifi
15	96.5	5.1	1085	1 CARB_HELPY	O25577 helicobacte
16	93.5	5.0	702	1 HELS_AERPE	Q9yfq8 aeropyrum p
17	93	4.9	312	1 MDH_BACIS	Q59202 bacillus is
18	93	4.9	1067	1 LONM_SCHPO	Q09769 schizosacch
19	93	4.9	4829	1 BIR6_HUMAN	Q9nr09 homo sapien
20	92.5	4.9	349	1 IDI2_SYNY3	P74287 synecocyst
21	92.5	4.9	723	1 RRPO_TNVA	P22958 tobacco nec
22	92	4.9	428	1 Y047_METJA	Q60355 methanococc
23	92	4.9	1570	1 P3K1_DICTDI	P54673 dictyosteli
24	91.5	4.9	682	1 HS7M_SOLTU	Q08276 solanum tub
25	91.5	4.9	1018	1 SYI_ARCFU	O29622 archaeoglob
26	91	4.8	556	1 ADEC_METJA	Q58854 methanococc
27	90	4.8	862	1 HELX_METTH	O27830 methanobact
28	90	4.8	871	1 SYA_AQUIB	Q9xdm3 aquifex pyr
29	89.5	4.8	484	1 GLGA_VIBCH	Q9krb6 vibrio chol
30	89	4.7	344	1 CHEB_THEMEA	Q9wyn9 thermotoga
31	89	4.7	447	1 YG99_PYRHO	O59362 pyrococcus
32	89	4.7	571	1 ILVI_BUCAI	P57321 buchnera ap
33	89	4.7	658	1 REP_BUCAP	O51889 buchnera ap

34	89	4.7	669	1 COTE_HUMAN	P81408 homo sapien
35	89	4.7	1125	1 MFD_BORBU	O51568 borrelia bu
36	88.5	4.7	223	1 GTAL_RABIT	Q08863 oryctolagus
37	88.5	4.7	467	1 UDPG_MUSAC	Q9sdx3 musa acumin
38	88.5	4.7	630	1 YZ35_METJA	Q60291 methanococc
39	88	4.7	337	1 INT2_ECOLI	P09999 escherichia
40	88	4.7	378	1 RPSD_CLOAB	P33656 clostridium
41	87.5	4.7	297	1 PUR7_MYCTU	Q59566 mycobacteri
42	87.5	4.7	393	1 PSMR_SULSO	Q980ml sulfolobus
43	87.5	4.7	393	1 RPN9_YEAST	Q04062 saccharomyc
44	87.5	4.7	476	1 UDPG_SOLTU	P19595 solanum tub
45	87.5	4.7	496	1 C72C_ARATH	Q9zu07 arabidopsis

ALIGNMENTS

RESULT 1

AMYA_PYRHO	STANDARD;	PRT;	633 AA.
ID	AMYA_PYRHO		
AC	O57932;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Alpha-amylase (EC 3.2.1.1).		
GN	AMYA OR PHO193.		
OS	Pyrococcus horikoshii.		
OC	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;		
OC	Pyrococcus.		
OX	NCBI_TaxID=53953;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=OT3;		
RX	MEDLINE=98344137; PubMed=9679194;		
RA	Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,		
RA	Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,		
RA	Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,		
RA	Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,		
RA	Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,		
RA	Masuchi Y., Shizuya H., Kikuchi H.;		
RT	"Complete sequence and gene organization of the genome of a hyper-		
RT	thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";		
RL	DNA Res. 5:55-76(1998).		
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic		
CC	linkages in oligosaccharides and polysaccharides.		
CC	-!- PATHWAY: POLYSACCHARIDE DEGRADATION.		
CC	-!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AP000001; BAA29262.1; -		
DR	InterPro; IPR004300; Glyco_hydro_57.		
DR	Pfam; PF03065; Glyco_hydro_57; 1.		
KW	Hydrolase; Glycosidase; Carbohydrate metabolism; Complete proteome.		
SQ	SEQUENCE 633 AA; 75017 MW; 9DCBCBB2FE191501 CRC64;		

Query Match 8.6%; Score 161.5; DB 1; Length 633;
Best Local Similarity 23.6%; Pred. No. 6e-05;
Matches 96; Conservative 65; Mismatches 134; Indels 111; Gaps 26;

QY	8	GNLQVAEIPKSEIPKVIKAYIPVETLIKEIP---FGLNITGYTLKFLPK---DIIDL 61
		:: : : : : :
Db	17	GNFEW-----IKRAYEKAYRPFLETL--EYYPNMKVAVHISGLVLELNRNPEIDL 68
QY	62	VKGGIASDLIEIGTSYTHAILPLLP-LSRVEAQVQRDRVEKEL-FEVS PKGFWLPELA 119
		: : : : : : : : : : : : : : :
Db	69	LKSLIKKQQLVWAGFYEPILVAIPEEDRVE-QIKLSKGWARKMGYEA--RGLWLTERV 125


```
QY 120 YDPIIPAILKONGVEYLFADGEAMLFSAHLNSA-IKPIKPLYPHLIKAKREKFRY---- 174
Db 126 WEPELVKTLREAGIEYILD-----DIHFSAGLSKEELFWPYTENGGEAIVFPIDE 179
QY 175 -ISYLLGLRELKAIKLVF-----EGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLM 227
Db 180 KLRYLIPRPVNETLEVYLSLADEDESKVAV-FHDDGEKFGAWPGTHLVY----- 229
QY 228 NPKKVASWI-----EDKDNILLYGTDIEFIG-----YRDIAGY-RMSVEGL-- 267
Db 230 -----ERGWLKEFFDRISDDKINLMYS-----EYLSKFRPKGLVYLPFIASFEMSEWSLPA 282
QY 268 -----LEVIDELNSELCLPSELKHSGRELYLRTSSWAPDKSLRIWREDEGN---ARLNM 318
Db 283 RQAKLFFEFIKKL-KELNLFKYR-----IFVRGGIW---KNF-LYKYPEGNYMHKRMMLM 332
QY 319 LSYNMRGELAFLAENSDARGWEPLPERRLDARAIYND--WRGENG 362
Db 333 LSKLLRNN-----PTARIFVLRAQCNDAYWHGVFG 362
```

RESULT 2

```
AMY1_DICTH STANDARD; PRT; 685 AA.
ID AMY1_DICTH AC P09961;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Alpha-amylase 1 (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase).
GN AMYA.
OS Dictyoglomus thermophilum.
OC Bacteria; Dictyoglomus group; Dictyoglomus.
OX NCBI_TaxID=14;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=H-6-12;
RX MEDLINE=88225097; PubMed=2453362;
RA Fukusumi S., Kamizono A., Horinouchi S., Beppu T.;
RT "Cloning and nucleotide sequence of a heat-stable amylase gene from
an anaerobic thermophile, Dictyoglomus thermophilum.";
RL Eur. J. Biochem. 174:15-21(1988).
CC -!- FUNCTION: THIS AMYLASE IS A HIGHLY LIQUEFYING-TYPE; OLIGOMERS
APPEARED AT THE BEGINNING OF INCUBATION, FOLLOWED BY A GRADED
DECREASE IN THE AMOUNTS OF MALTOTRIOSE, MALTOSE AND GLUCOSE IN
PROLONGED INCUBATION. IT IS HIGHLY HEAT-STABLE; THE OPTIMUM
TEMPERATURE FOR ITS ACTIVITY WAS FOUND TO BE AROUND 90 DEGREES
CELSIUS, BUT DECREASED TO ONLY 70% ACTIVITY AFTER 1 HOUR. NO
DECREASE OF ACTIVITY WAS OBSERVED WITHIN THE SAME TIME AT 80
DEGREES CELSIUS.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
linkages in oligosaccharides and polysaccharides.
CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

Query Match

8.5%; Score 159.5; DB 1; Length 685;

```
Best Local Similarity 21.0%; Pred. No. 9.6e-05;
Matches 85; Conservative 64; Mismatches 149; Indels 107; Gaps 19;
QY 8 GNLQYAEIPKSEIPKVIKAYIPVIETLIKE-EIPFGLNITGYTLKFLPK---DIIDLVK 63
Db 17 GNDFDV-----IERAYMSYKPLINFFFKHPDPINVFHSGFLLWLEKNHPEYFEKLK 70
QY 64 GGIASDLIEIIGTSYTHAILPLPLSRVEAQVORDREVKEELFEVSPKGFWLPDAYDPI 123
Db 71 IMAERGOIEFVSGGFYEPIIPIDKDKVQIQIKLNKYYIDKFGQTPKGMWLAERVWEPH 130
QY 124 IPAILKDNNGEYELFADGEAMLFSAHLNS-----AIKPIKPLYPHLIK-A 166
Db 131 LVKYIAEAGIEYVVVD-DAHFFSVGLKEEDLFGYYLMEEQGYKLAVFPISMKLYLIPFA 189
QY 167 QREKRFRYISYLLGLRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPL 226
Db 190 DPEETITYLDKFFAS--EDKSKIALLFD-----DGEKFGLPDPTYRTVY----- 230
QY 227 MNPKKVASWIED-----KDNILL-----YGTDFIEFIG--YRDIAGYRMSVEGLL--- 268
Db 231 -----EEGWLETFVSKIKENFLLVTPVNLTYTMQRVKPKGRIYLPNTASYREMMEWVLFPE 285
QY 269 -----EVIDELNSELCLPSELKHSGRELYLRTSSWAPDKSLRIWRE-----DEGNARLN 317
Db 286 AQKELEELVEKLKTE-----NLW---DKFSPYVKGFGFWRNFLAKYDESN----- 326
QY 318 MLSYNMRGELAFLAENSDARGWEPLPERRLDARAIYND--WRG 359
Db 327 ----HMQKKMLYVWKVKVQDSPNEEVKEKAMEEVFQGDANDAYWHG 367
```

RESULT 3

```
AMYA_METJA STANDARD; PRT; 467 AA.
ID AMYA_METJA AC Q59006;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative alpha-amylase (EC 3.2.1.1).
GN MJ1611.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
linkages in oligosaccharides and polysaccharides.
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

TIGR; MJ1611; -
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57; 1.
KW Hypothetical protein; Hydrolase; Glycosidase; Carbohydrate metabolism;
SQ SEQUENCE 467 AA; 55558 MW; 40A6B1CDDDD4D967E CRC64;
Query Match 7.8%; Score 146.5; DB 1; Length 467;
Best Local Similarity 23.1%; Pred. No. 0.00058;
Matches 87; Conservative 66; Mismatches 129; Indels 95; Gaps 24;
QY 14 EIPKSEIPKVIKAYIPVETLIK--EIPFGLN--ITG----YTLKFLPKD-IIDLKVG 64
Db 34 KLNKEVFNKANKCYIPTNELILEIDYDFKVNYSITGVFVEQALEF--NDYVLDLFD 91
QY 65 GIASDLIEIGTSYTHAILPLPL-SRVEAQVQDRREVKEELFEVSPKGFWLPELAYDPI 123
Db 92 LVKTGNVELIAETYYHSLTSLFTEDEFIEDIEMHRKMYKEIFGFKAKVFRNTELYNNR 151
QY 124 IPAILKDNQVEYLFADG-EAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRYIS---YLL 179
Db 152 IAKIAKDLGPKAIFTEGIEKIL-----GWRSPNYLYQSPDGMKI 190
QY 180 GLRELKAIKLVFEGKVTLKAVKDIEAVP-----VWAVNTAVMLGIGRLPLMNPKKV 232
Db 191 LLRNRLSDDIGFR-----FSARDWQYPLTADKYAIWLASTPGEVINI---YMDYETF 241
QY 233 AS--WIEDKNILLYGTDI-EFIGYR--DIAGY-RMSVEGLEVIDELNSELCLPSLKH 286
Db 242 GEHWKE-----TGIFEFLRYLPIEIAKHEHLEVVNVSEVVDRLPR----- 283
QY 287 SGRELYLR---TSSWA-PDKSLRIWREDEGNARLNMLSYNMRGELA-FLAENSARGWEP 341
Db 284 --GEIYVHEPATISWADTERDVSANL---GN-KMQRISFEKLDIGKFIKENSANL 333
QY 342 LPERRLDARAIYNDWR 358
Db 334 ---KKLNKFDEIYKMYK 347
RESULT 4
AMYA_PYRAB STANDARD; PRT; 655 AA.
AC Q9V298;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-amylase (EC 3.2.1.1).
GN AMYA OR PAB0118.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic linkages in oligosaccharides and polysaccharides.
CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AJ248283; CAB49100.1; -
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57; 1.
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Complete proteome.
SQ SEQUENCE 655 AA; 77296 MW; 7F6F920B1A00EECE CRC64;
Query Match 7.5%; Score 141.5; DB 1; Length 655;
Best Local Similarity 21.9%; Pred. No. 0.0022;
Matches 88; Conservative 54; Mismatches 130; Indels 129; Gaps 21;
QY 25 EKAYIPVETLIKKEIP---FGLNITGYTLKFLPK---DIIDLKVGSIASDLIEIGTSY 78
Db 28 EKAYRPFLEIL--EYPNMKVAIHISGILVLEENKPDYIDLKSLVRKQGVIVVAGF 85
QY 79 THAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPELAYDPIIPAILKDNQVEYLF 138
Db 86 YEPVLAAPKEDRLQIYLLKEWAKKI-GYDAKGLWLTERVWQPELVKTLREAGIEYVV 144
QY 139 DGEAMLFSAHLNSA-IKPIKPLYPHLIKAQRE-----KFRY-----ISYLL 179
Db 145 D-----DYHFMASGLSKDQLFWPYTTEGGGEVITVPIDEKRLYLIPERFVVKVISYLH 198
QY 180 GLRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDK 239
Db 199 SLASEDESKVAVFH-----DDGEKFGIW-----PM-----TYEWVYEK 231
QY 240 D-----NILLYGTDI-----FIGYRDIAGY-----RMSVEG 266
Db 232 GWLREFFDRVSSDEAINIMLYSEYLOKFKPKGLVLPPIASYFEMSEWSLPAQAKLFEV- 290
QY 267 LLEVIDELNSELCLPSLKHSGRELYLRTSSWAPDKSLRIWREDEGN---ARLNMLSYNM 323
Db 291 FVEKIKELN-----MFERYRVFVRGGIW---KNF-FYKYPEANYMHKRMMLSLRL 337
QY 324 RGEALFLAENSARGWEPLPERRLDARAIYND--WRGENG 362
Db 338 RDN-----PSARRFVLRAQCNDAYWHGVFG 362
RESULT 5
AMYA_PYRFU STANDARD; PRT; 648 AA.
AC P49067;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alpha-amylase (EC 3.2.1.1).
GN AMYA OR PF0272.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC MEDLINE=94043280; PubMed=8226990;
RA Laderman K.A., Asada K., Uemori T., Mukai H., Taguchi Y., Kato I., Anfinsen C.B.;
RT "Alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus furiosus. Cloning and sequencing of the gene and expression in Escherichia coli.";
RL J. Biol. Chem. 268:24402-24407(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RN Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RP CHARACTERIZATION.
RX MEDLINE=94043279; PubMed=8226989;
RA Laderman K.A., Davis B.R., Krutzsch H.C., Lewis M.S., Griko Y.V., Privalov P.L., Anfinsen C.B.;
RT "The purification and characterization of an extremely thermostable


```
CC CC -!- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; U05257; AAA96343.1; -.
CC DR EMBL; AF008220; AAC00347.1; -.
CC DR EMBL; 299118; CAB14872.1; -.
CC DR HSSP; Q27743; ICET.
CC DR Subtilisin; BGI1386; cith.
CC DR InterPro; IPR001557; L_LDH.
CC DR InterPro; IPR001252; Mdh.
CC DR InterPro; IPR000205; NAD_binding.
CC DR InterPro; IPR001236; ldh.
CC DR Pfam; PF00056; ldh; 1.
CC DR Pfam; PF02866; ldh_C; 1.
CC DR PRINTS; PR00086; LLDHNRGNASE.
CC DR PROSITE; PS00068; MDH; FALSE_NEG.
CC KW Oxidoreductase; Tricarboxylic acid cycle; NAD; Complete proteome.
CC FT INIT_MET 0
CC FT ACT_SITE 152 152 PROTON-RELAY (BY SIMILARITY).
CC FT BINDING 155 155 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
CC FT ACT_SITE 179 179 PROTON-RELAY (BY SIMILARITY).
CC SQ SEQUENCE 311 AA; 33512 MW; 4752B7BF852FB09 CRC64;
```

```
Query Match 5.8%; Score 109; DB 1; Length 311;
Best Local Similarity 23.2%; Pred. No. 0.27;
Matches 44; Conservative 32; Mismatches 58; Indels 56; Gaps 9;
QY 10 LQYAEIPKSEIPK-----VIEKAYIPVIETLKEEIPFGLNITGYTLKFLPKDIIDL 61
Db 129 MTYAVYKESGFPKERVIGSGVLDTR---FRTEFAEE---LNLS-----VKDVTGF 174
QY 62 VKGGIASDLIEIIGTSYTHAI--LPLPLSRVDAQVQDRREVKEELFEVSPKG--FWLPE 117
Db 175 VLGGHGDMMVPLVRSYAGGIPLTLIPKERTDAIVERTKGGGEIVNLLNGSAYAPA 234
QY 118 LAYDPIIPAILKDN-----GYEYLF-----ADGEAMLF-----A 147
Db 235 ASLTEMVEAILKDRRLVPTIAYLEGEYGEIYLGVTIVGGNGLEQIIIELELTDYERA 294
QY 148 HLNSAIKPIK 157
Db 295 QLNKSVEVK 304
```

```
RESULT 9
MDH_BACHD STANDARD; PRT; 314 AA.
AC Q9K849;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
GN MDH OR CITH OR BH3158.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RT Nucleic Acids Res. 28:4317-4331(2000).
```

```
CC CC -!- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC CC -!- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; AP001517; BAB06877.1; -.
CC DR HSSP; Q27743; ICEQ.
CC DR InterPro; IPR001557; L_LDH.
CC DR InterPro; IPR001252; Mdh.
CC DR InterPro; IPR000205; NAD_binding.
CC DR InterPro; IPR001236; ldh.
CC DR Pfam; PF00056; ldh; 1.
CC DR Pfam; PF02866; ldh_C; 1.
CC DR PRINTS; PR00086; LLDHNRGNASE.
CC DR PROSITE; PS00068; MDH; FALSE_NEG.
CC KW Oxidoreductase; Tricarboxylic acid cycle; NAD; Complete proteome.
CC FT ACT_SITE 153 153 PROTON-RELAY (BY SIMILARITY).
CC FT BINDING 156 156 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
CC FT ACT_SITE 180 180 PROTON-RELAY (BY SIMILARITY).
CC SQ SEQUENCE 314 AA; 33680 MW; 02D132F11E3B8E34 CRC64;
```

Query Match 5.3%; Score 100.5; DB 1; Length 314;

Best Local Similarity 22.6%; Pred. No. 1.3;

```
Matches 49; Conservative 33; Mismatches 74; Indels 61; Gaps 9;
QY 10 LQYAEIPKSEIPK-----VIEKAYIPVIETLKEEIPFGL-NITGYTLKFLPKDIID 60
Db 130 MTYVYKESGFPKRVIGSGVLDTR---FRTEFAEQLNLSVEDITGFVL----- 177
QY 61 LVKGGIASDLIEIIGTSYTHAI--LPLPLSRVDAQVQDRREVKEELFEVSPKG--FWLP 116
Db 178 ---GGHGDMMVPLIRYSYAGGIPLKLLPQERIDAIVERTKGGGEIVLNGSAYAP 234
QY 117 ELAYDPIIPAILKDNNGYEVLFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRY-- 174
Db 235 AASLAEMVEAILKDK-----KRVLTIAYL-----EGEYGYED 267
QY 175 -----ISYLLGLRELKAIKLVFEGKVTLKAVKDIEAV 207
Db 268 IYGVPTILGGDGIKGVIELDLTDEKATFAKSIESV 304
```

RESULT 10

```
NOEK_RHISN STANDARD; PRT; 474 AA.
AC P55356;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphomannomutase (EC 5.4.2.8) (PMM).
GN NOEK OR Y4AI.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -!- CATALYTIC ACTIVITY: D-mannose 1-phosphate = D-mannose 6-phosphate.
CC -!- PATHWAY: BIOSYNTHESIS OF THE FUCOSE MOIETY OF THE NOD FACTOR.
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOHEXOSE MUTASES FAMILY.
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000064; AAB91606.1; -
DR InterPro: IPR001485; PG/PMM_mutase.
DR Pfam; PF00408; PGM_PMM; 1.
DR Pfam; PF02878; PGM_PMM_I; 1.
DR Pfam; PF02879; PGM_PMM_II; 1.
DR Pfam; PF02880; PGM_PMM_III; 1.
DR PROSITE; PS00710; PGM_PMM; 1.
KW Isomerase; Phosphorylation; Nodulation; plasmid.
FT ACT_SITE 102 102 FORMS THE PHOSPHOSERINE INTERMEDIATE
FT (BY SIMILARITY).
SQ SEQUENCE 474 AA; 49670 MW; 6965310CEA96A22B CRC64;

Query Match 5.3%; Score 99; DB 1; Length 474;
Best Local Similarity 20.3%; Pred. No. 2.9;
Matches 80; Conservative 59; Mismatches 128; Indels 128; Gaps 16;

QY 15 IPKSEIPKVEKA-----YIPVIETLIKEEIPF 42
Db 114 LPDGEINKADEQAITALAEQLSADADATRVCEGRGADHSSEATDFYIQRYETLLPKSGLK 173
QY 43 GLNITGYTLKFLPKDIIDLKVGGIASDLIEIIGTSYTHAILPLLSRVEAQVQDREVK 102
Db 174 GLKIGLQHSSVARDILTITILEGHANVV-----PVGRSEVFIPVDTEA- 217
QY 103 EELFEVSPK--GFWLPELAYDPIIPA-----ILKNDNGVEYLFAD-----GEAML 144
Db 218 --ISAATCKMLAAWAKEFAFDAIVSSDADADRPLLTDETGTPLRGDLLGLICARLLEAKL 275
QY 145 FSAHL--NSAIKPIKPL-----YPHLIKAQREKFRFYISYLLGLRELKAIKLVPE- 193
Db 276 IATPITSNGIEAASGVEVTRVGSPIVIAAMTEAVARGKQVGMGF-EANGGVMLGSNF 334
QY 194 --GKVTLKAVKDIEAV-PVWVAVNTAV-----MLGIGRLPLMNPKNKVASWIEDKNI 242
Db 335 SFGGASLPALPTRDCVLPPIIAALHMAVEAKTPLSGIVAMHRLPVALSGRINIYPPDRSDA 394
QY 243 L-----LYGTDIEFIGYRDIAGRMSVEGLLEVIDELNSELCLPSELKHSGR 289
Db 395 LVAFKASKANVSHLFSRIGRVAGTDDVDGLRLTFEG-----GR 433
QY 290 ELYLRTSSWAPKSLRIWRE-DEGNARLNMLSYNM 323
Db 434 ILHIRPSGNAPE--LRCYVEADDPDAAEHLAQGL 466

RESULT 11
RNR_AQUAE STANDARD; PRT; 705 AA.
AC O67834;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonuclease R (EC 3.1.-.-) (RNase R) (VacB protein homolog).
GN RNR OR VACB OR AQ_2046.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aufay M., Huber R.,

RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL Nature 392:353-358(1998).
CC -!- FUNCTION: 3'-5'EXORIBONUCLEASE THAT PARTICIPATES IN AN ESSENTIAL
CC CELL FUNCTION. ACTS NONSPECIFICALLY ON POLY(A), POLY(U) AND
CC RIBOSOMAL RNAs (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.
CC -!- SIMILARITY: CONTAINS 1 S1 MOTIF DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000769; AAC07792.1; -
DR InterPro: IPR004476; 3_prime_RNase.
DR InterPro: IPR002059; Cold_shock.
DR InterPro: IPR001900; Ribonuclease_II.
DR InterPro: IPR003029; S1.
DR Pfam; PF00575; S1; 1.
DR Pfam; PF00773; RNB; 1.
DR SMART; SM00357; CSP; 1.
DR SMART; SM00316; S1; 2.
DR TIGRFAMs; TIGR00358; 3_prime_RNase; 1.
DR PROSITE; PS01175; RIBONUCLEASE_II; 1.
DR PROSITE; PS0126; S1; 1.
KW Hydrolase; Exonuclease; Nuclease; RNA-binding; Complete proteome.
FT DOMAIN 615 696 S1 MOTIF.
SQ SEQUENCE 705 AA; 81537 MW; 2C08587229D1B05E CRC64;

Query Match 5.2%; Score 98.5; DB 1; Length 705;
Best Local Similarity 16.3%; Pred. No. 5.4;
Matches 76; Conservative 80; Mismatches 114; Indels 195; Gaps 22;

QY 5 VFHGNLQYAEIP---KSE--IPKVIEKAYIPVIETLIKEE-----IPFGLNITGYTLKF 53
Db 100 VFHGDVVKAKVTEFGKKEVRIIKVLKRAKKDIVAKVVFDEQCYVVPDEN-AHRRILL.158
QY 54 LPKDIIDLKVGKGGIA-----SDLIEIG-----TSYT 79
Db 159 SKKDCQKLKEGEVVVLKITQFPPTKKSPARGKIVIEVLGNPKKFKFIAIDVIIRKYNLPTSY 218
QY 80 HAIL---PLLPLSRVEAQVQDREVKEEL-----FEVSPKGFW----- 114
Db 219 EKVIKEVEAIPPEEIEIKRRDLREQLCFTIDPEKAGDFDDAVAIELTPEGYKLYVH 278
QY 115 LPELAYDPIIPAILKD-----NGYEYLFADGEAMLSAHLNSAIKPIKPLYPHLI 164
Db 279 IADVSY-----YVREGTETDKAYKRGFTYTFPPDRALHMLPEKLSAKLSLRP----- 326
QY 165 KAQREKRFYISYLLGLRELKRAIKLVFEGKVTLLKAVKDIEAV---PVWVAVNTAVMLGI 221
Db 327 ---NEDKLAF-----TVEMVFDESGNLKAYDIYESVIRSKARLTYNEALALIV 371
QY 222 G-----RLP-LMNPKKVAS-----W----- 235
Db 372 GDPALEKKFPNLVEPLRMETLYRLSRKRWEMSGSIDFDLPEAEVIVDEYGEPTAIYPYE 431
QY 236 -----IEDKDNILLYGTDFIEFIGYRDIAGYRM-----SVEGLLEVIDELNS 276
Db 432 RHVAHRIIEHFMIANETVALH---LEHAGYPCPL--YRVHEPDEEKVENLLEILEGLY 486
QY 277 ELCLPSELKHSGRELYLRTSSWAPDKSLRIWREDEGNARLNMLSY 321
Db 487 KVKRPHE-----YTPKFFOKILEDFFGRPEENLVRF 517

RESULT 12
LON2_BORBU

ID LON2_BORBU STANDARD; PRT; 813 AA.
AC O51558;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent protease La homolog (EC 3.4.21.-).
GN BB0613.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Wathley L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi.";
RL Nature 390:580-586(1997).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001162; AAC66962.1; -
CC MEROPS; S16.UPW; -
CC TIGR; BB0613; -
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003959; AAA_ATPase_cent.
CC InterPro; IPR003111; LON.
CC InterPro; IPR001984; Lon_endopep.
CC InterPro; IPR004815; Lon_fam.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF02190; LON; 1.
CC PRINTS; PR00830; ENDOLAPTASE.
CC SMART; SM00382; AAA; 1.
CC SMART; SM00464; LON; 1.
CC TIGRFAMS; TIGR00763; lon; 1.
CC PROSITE; PS01046; LON_SER; 1.
KW Hydrolase; Serine protease; ATP-binding; Complete proteome.
FT NP_BIND 369 376 ATP (POTENTIAL).
FT ACT_SITE 719 719 BY SIMILARITY.
SQ SEQUENCE 813 AA; 92312 MW; 5129AA1498C5D0F6 CRC64;
Query Match 5.2%; Score 98.5; DB 1; Length 813;
Best Local Similarity 21.7%; Pred. No. 6.6;
Matches 72; Conservative 52; Mismatches 113; Indels 95; Gaps 18;
QY 12 YAEIPKSE-----IPKVEIKAYIPVETLIKEEIPFGLNITGYT----- 50
Db 507 YSIEKLEIAKIFLIPSIIKESFLDKYIRIEDDVIFNL-IRNYTMESGVRGLKRVLTNL 565
QY 51 LKFLPKDII-----DLVKGGI--ASDLIEIIGTSYTH-----AILPLPLSRVEAQVQR 97
Db 566 IRLVRELLYEYSKQDIKGNFYSPSSLIHGNSLFTHPDIPGIYKIININNYV-- 623
QY 98 DREVKEELFEVSPKGF-----W-----LP-----ELAYDPIIPAILKDNNGVEY 135
Db 624 DTEADNLDLIKIDSSGVFVGLAWTNYGGTVLPVEATKFEKKGDIIITGSLGAIMKESAQ-- 681
QY 136 LFADGEAMLFSAHLNSAIKPIKPLPHLIKAQREKFRFYISVLLG-----LRELKAIK 189

Db 682 -LAYSIVKTYSSKLNFDVKESPEIHLHFPPEGATPKDGPAGITATAIASILSDKKVPLD 740
QY 190 LVFEGKVTILKAVKDIEAVPVW---VAVNTAVMLGIGR--LPLMNPKKVASWIED-KDNIL 243
Db 741 LAMTGEVTLKGF-----VLPVGGIKEKVLAAAYRNGISKVILPKDNKKDYSKLPEEVKDN1- 795
QY 244 LYGTDIEFIGYRDIAGYRMSVEGLLEVIDELN 275
Db 796 ----DVKE-----VSSLEEVFDYLN 811
RESULT 13
MDH_BACTC
ID MDH_BACTC STANDARD; PRT; 312 AA.
AC Q9X4K8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
GN MDH.
OS Bacillus thermodenitrificans.
OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=33940;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=00462;
RA Williams R.A.D., Welch S.G., Alawadhi S.A.;
RT "Properties and primary structure of a thermostable L-malate
RT dehydrogenase from 'Bacillus thermodenitrificans'";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -!- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF114423; AAD28555.1; -
CC HSSP; Q27743; ICET.
CC InterPro; IPR001557; L_LDH.
CC InterPro; IPR001252; Mdh.
CC InterPro; IPR000205; NAD_binding.
CC InterPro; IPR001236; ldh.
CC Pfam; PF00056; ldh; 1.
CC Pfam; PF02866; ldh_C; 1.
CC PRINTS; PR00086; LLDHNRGNASE.
CC PROSITE; PS00068; MDH; FALSE_NEG.
KW Oxidoreductase; Tricarboxylic acid cycle; NAD.
FT ACT_SITE 153 153 PROTON-RELAY (BY SIMILARITY).
FT BINDING 156 156 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
FT ACT_SITE 180 180 PROTON-RELAY (BY SIMILARITY).
SQ SEQUENCE 312 AA; 33682 MW; 68C691C462EFF452 CRC64;

Query Match 5.2%; Score 97.5; DB 1; Length 312;
Best Local Similarity 21.8%; Pred. No. 2.1;
Matches 47; Conservative 34; Mismatches 76; Indels 59; Gaps 9;
QY 10 LOYAEIPKSEIPK-----VIEKAYIPVETLIKEEIPFGLNITGYTLKFLPKDIIDL 61
Db 130 MSYTVFKESGFKNRVIGSGVLDTR---FRTEVQAQ-----LNIS-----VKDVTGF 175
QY 62 VKGGIASDLIEIIGTSYTHAI--LPLPLSRVEAQVQRDREVKEELEVSPKG--FWLPE 117
Db 176 VLGGHGDMMVPLVRYSYAGGIPLKIPKDLDAIVERTRKGGGEIVNLLNGSAYAPA 235
QY 118 LAYDPIIPAILKDNNGVEYLFADGEAMLFSAHLNSAIKPIKPLPHLIKAQREKFRFY--- 174
Db 236 ASLAEMVEAIVKDO-----RRILPAITYL-----EGEYGYEGI 268

Search completed: November 18, 2002, 07:25:29
Job time : 18 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 07:24:24 ; Search time 22 Seconds
(without alignments)
1590.588 Million cell updates/sec

Title: US-09-619-032-4
Perfect score: 1880
Sequence: 1 MRALVFHGNLQYAEIPKSEI.....RRLDFAFRAYNDWRGENGP 364

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1506	80.1	364	2 E71144	hypothetical prote
2	167	8.9	529	2 AC2112	hypothetical prote
3	161.5	8.6	529	2 S76831	hypothetical prote
4	161.5	8.6	633	2 G71241	probable alpha-amylase A (EC 3.2.1)
5	159.5	8.5	686	1 ALDYAT	uncharacterized co
6	151	8.0	527	2 F97197	alpha-amylase (EC
7	146.5	7.8	467	2 B64501	alpha-amylase (or
8	141.5	7.5	655	2 E75206	alpha-amylase (EC
9	140.5	7.5	649	2 A49512	hypothetical prote
10	129.5	6.9	560	2 D71011	hypothetical prote
11	127	6.8	923	2 G83826	amylopullulanase P
12	126	6.7	1362	2 A75207	hypothetical prote
13	123.5	6.6	744	2 AC1970	hypothetical prote
14	119	6.3	602	2 C75120	methanol dehydroge
15	115.5	6.1	324	1 B69553	conserved hypothet
16	111	5.9	526	2 D71334	alpha amylase (imp
17	109.5	5.8	447	2 E90250	malate dehydrogena
18	109	5.8	312	2 I40383	hypothetical prote
19	109	5.8	1069	2 AF1930	conserved hypothet
20	107.5	5.7	902	2 E90270	hypothetical prote
21	104	5.5	526	2 B70859	serine/threonine k
22	104	5.5	546	2 AI2045	hypothetical prote
23	103	5.5	704	2 G96587	pyruvate carboxyla
24	103	5.5	1150	2 G89881	probable serine/th
25	101.5	5.4	684	2 T47694	Bir repeat contain
26	101	5.4	4845	2 T31067	probable xylanase/
27	100.5	5.3	266	2 G96943	malate dehydrogena
28	100.5	5.3	314	2 F84044	probable alpha-amylase
29	100.5	5.3	619	2 D71361	

30	99.5	5.3	1181	2 T01799	hypothetical prote
31	99	5.3	503	2 T19319	hypothetical prote
32	99	5.3	636	2 S70877	hypothetical prote
33	99	5.3	815	2 G97266	mannose-1-phosphat
34	98.5	5.2	705	2 F70475	vacB protein (ribo
35	98.5	5.2	725	2 T44992	translation elonga
36	98.5	5.2	813	1 D70176	endopeptidase La (
37	98	5.2	510	2 B70368	L-aspartate oxidas
38	97.5	5.2	588	2 AE1926	two-component resp
39	97	5.2	764	1 H70414	conserved hypothet
40	96.5	5.1	627	2 D84531	probable salt-indu
41	96.5	5.1	1085	2 G64634	carbamoyl-phosphat
42	96	5.1	530	2 F82271	transglycosylase,
43	95.5	5.1	1134	2 D75014	hypothetical prote
44	95	5.1	394	2 F75006	hypothetical prote
45	94.5	5.0	453	2 D84018	two-component resp

ALIGNMENTS

RESULT 1
E71144

hypothetical protein PH0368 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 08-Sep-2000

C:Accession: E71144

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophili

A:Reference number: A71000; MUID:98344137; PMID:9679194

A:Accession: E71144

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-364 <KAW>

A:Cross-references: GB:AP000002; NID:g3236129; PIDN:BAA29442.1; PID:d1030385; PID:g32

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH0368

C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0368

Query Match 80.1%; Score 1506; DB 2; Length 364;
Best Local Similarity 76.6%; Pred. No. 2e-106;
Matches 278; Conservative 47; Mismatches 38; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIEKAYIPVIETLIKEIPFGINITGYTLKFLPKDIID 60

Db 1 MRALIFHGNLQYAEIPKHEISKVIEKSYFPTISELIKREIPFGINITGYSLQFLPQELIH 60

QY 61 LVKGGIASDLIEIGTSYTHAILPLPLSRVEAQVQDRKVEELFEVSPKGFWLPELAY 120

Db 61 LIKEGIESELIEILGTSYTHAILPLTLSRIEAQIKRDREIKEIFEVSPGGFWLPELAY 120

QY 121 DPTIIPAILKDNQYGYEYLFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRYISYLLG 180

Db 121 DPTIIPAILRDNEYEYLFADGEAMLFSAHLNSAIKSIKPLYPYLIKAQREGFVYLYLLG 180

QY 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240

Db 181 LRELKKAINLTFGGKVTLEAVKDIEAIPVWVSINIAIMLGAGRPPLMSPKRVANWIKGD 240

QY 241 NILLYGTDFIEGYRDIAGYRMSVEGLLEVIDELNSELCLPSLKHSGRELYLRTSSWAP 300

Db 241 EILLYGTDFIEFLGYRSIAGHKIAISGLMEVLNSELGELCLPRDRHNGRRLYLRTSSWAP 300

QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSADARGWEPLPERRLDFAFRAYNDWRGE 360

Db 301 DKSLRIWKEDEGNARLNMLTYCMDSGEFAFLAENSADARGWEPLPERRLDFAFKAIYKYWRNE 360

QY 361 NGE 363

||:


```

Db 191 DPEETIYLDKFAS--EDKSKIALFLD-----DGEKFLWPDTYRTVY----- 231
Qy 227 MNPKKVASWIED-----KDNILL-----YGTDIEFIG--YRDIAGYRMSVEGLL--- 268
Db 232 -----EEGWLETFVSKIKENFLLVTPVNLTYMQRVKPKGRIYLPQTASYREMMEWVLFPE 286
Qy 269 -----EVIDELNSELCLPSELKHSGRELYLRTSSWAPDKSLRIWRE-----DEGNARLN 317
Db 287 AQKELEELVEKLKTE-----NLW---DKFSPYVKGGFWRNFALAKYDESN----- 327
Qy 318 MLSYNNRGELAFLAENSDARGWEPLPERRL-DAFRAIYND--WRG 359
Db 328 ---HMQKKMLYVWKVKVQDSPNEEVKEKAMEVFQOQANDAYWHG 368

RESULT 6
F97197
Uncharacterized conserved protein, related to alpha-amylase/alpha-mannosidase CAC2414
S:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: F97197
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97197
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80369.1; PID:gl5025430; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824

```

```

C:Genetics:
A:Gene: CAC2414
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

Query Match      8.0%; Score 151; DB 2; Length 527;
Best Local Similarity 18.9%; Pred. No. 0.0012;
Matches 92; Conservative 63; Mismatches 119; Indels 212; Gaps 20;

QY 5 VFHGNLQYAEIPKSE-----IPKVIEKAYIPVIE--TLIKEEIPFG----- 43
   | | :: | :: : : : |||::| | : | |
Db 9 VLHSHMPFVRHPETRDSLEERWLFAMSECYIPLIEVYDNLKDNKIKFRMTMSITPPLMS 68

```

```

QY      177 YLLGLRELKAIKLVEFGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWI   236
           ||               |
Db      225 -----GTNT-----PIAAPSGVCAGF   240

QY      237 EDKNILLYGTDIEFTG-----YRDTAGYRMSVEGLLEVIDELNSELCLPSELKH   286
           | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      241 RDWDS--SYQVWSDFMGYPGDFNYREFYRDI-GFELPMEYIKPYINE--NGIRIDTGFKY   295

QY      287 -----SGRE-LYLRTSS-----WAPDKSLRIWREDEGN-ARLNM-----   318
           || : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      296 YKITNSGEKGIYNRENAMKKVWEHASHFASCRHDQINAAAAANMKPPITPCPYDTELYG   355

QY      319 -----LSYNMRGELAFLAENSARGWEPLPERRLDAFAIYNDW   357
           || : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      356 HWMEFCPEFINAEFIKRSADFOWTSYEIITPTTEYLKNNSMVOCCSSPS-----SSW   404

```

QY 358 RGENGE 363
||||:
Db 405 -GENGD 409

RESULT 7
B64501
alpha-amylase (EC 3.2.1.1) - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: B64501
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: B64501
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-467 <BUL>
A;Cross-references: GB:U67601; GB:L77117; NID:g2826439; PIDN:AAB99631.1; PID:g1592212; T
C;Genetics:
A;Map position: FOR1586396-1587799
C;Function:
A;Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A;Pathway: glycogen/starch degradation
C;Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 7.8%; Score 146.5; DB 2; Length 467;
Best Local Similarity 23.1%; Pred. No. 0.0023;
Matches 87; Conservative 66; Mismatches 129; Indels 95; Gaps 24;
QY 14 EIPKSEIPKVIKAYIPVIETLIK--EEIPFGLN--ITG-----YTLKFLPKD-IIDLKVG 64
Db 34 KLNKEVFNKANKYIPTNELILELIDEDYDFKNYSITGVFVEQALEF--NDYVLDLFKD 91
QY 65 GIASDLIEIGTSYTHAILPLPL-SRVEAQVQRDREVKEELFEVSPKGFWLPPELAYDPI 123
Db 92 LVKTGNVELIAETYYHSLTSLFTEDEFIEDIEMHRKMYKEIFGFKAKVFRNTELYNNR 151
QY 124 IPAILKDNQYEYLFADG-EAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRYIS---YLL 179
Db 152 IAKIADLGFKAIFTEGIEKIL-----GWRSPNYLYQSPDGMKI 190
QY 180 GLRELKRAIKLVFEGKVTLKAVKDIEAVP-----VWVAVNTAVMLGIGRLPLMNPKKV 232
Db 191 LLRNRYLSDDIGFR-----FSARDWDQYPLTADKYAIWLASTPGEVINI-----YMDYETF 241
QY 233 AS--WIEDKDNILLYGTDI-FIGYR--DIAGY-RMSVEGLLEVIDELNSELCLPSELKH 286
Db 242 GEHWKE-----TGIFEFLRYLPPIEIAKHEHLEVNVVSEVDRLEPR----- 283
QY 287 SGRELYLR---TSSWA-PDKSLRIWREDEGNARLNMLSYNMRGELA-FLAENSDARGWEP 341
Db 284 --GEIYVHEFATISWADTERDVSABL---GN-KMQRISFEKLDIGKFIKENS NKL----- 333
QY 342 LPERLDAFRAIYNDWR 358
Db 334 ---KKLNKFEIYKMYK 347

RESULT 8
E75206
alpha-amylase (or 4-alpha-glucanotransferase) PAB0118 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: E75206
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A;Reference number: A75001

A;Accession: E75206
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-655 <KAW>
A;Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49100.1; PID:g545
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: amYA; PAB0118
C;Superfamily: Dictyoglomus thermophilum amylase A

Query Match 7.5%; Score 141.5; DB 2; Length 655;
Best Local Similarity 21.9%; Pred. No. 0.0088;
Matches 88; Conservative 54; Mismatches 130; Indels 129; Gaps 21;
QY 25 EKAYIPVIETLIKEIP---FGLNITGYTLKFLPK---DIIDLKVGGIASDLIEIGTSY 78
Db 28 EKAYRPFLEIL--EYYPNMKVAIHISGILVLEENKPDYIDLKSLVRKQVEIVWAGF 85
QY 79 THAILPLPLSRVEAQVQRDREVKEELFEVSPKGFWLPPELAYDPIIPAILKDNQYEYLF 138
Db 86 YEPVLAAPKEDRLEOIYLLKEWAKKI-GYDAKGIMLTERVWQPELVKTLREAGIEYVVV 144
QY 139 DGEAMLFSAHLNSA-IKPIKPLYPHLIKAQRE-----KRFRY-----ISYLL 179
Db 145 D-----DYHFMSAGLSKQDLFWPYTYTEDGGEVITVFPIDEKRLYLIPRPVDPVKVISYLH 198
QY 180 GLRELKRAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDK 239
Db 199 SLASEDESKVAVFH-----DDGEKFGIW-----PM-----TYEWVYEK 231
QY 240 D-----NILLYGTDIE-----FIGYRDIAGY-----RMSVEG 266
Db 232 GWLREFFDRVSSDEAINIMLYSEYLOKFKPKGLVYLPIASYFEMSEWSLPAQAKLFVE- 290
QY 267 LLEVIDELNSELCLPSELKHSGRELYLRTSSWAPDKSLRIWREDEGN---ARLNMLSYNM 323
Db 291 FVEKLKELN-----MFERYRVFVRGGIW---KNF-FYKYPEANYMHKRMMLSLRL 337
QY 324 RGEALFLAENSDARGWEPLPERRLDARAIYND--WRGENG 362
Db 338 RDN-----PSARRFVLRAQCNDAYWHGVFG 362

RESULT 9
A49512
alpha-amylase (EC 3.2.1.1) - Pyrococcus furiosus
C;Species: Pyrococcus furiosus
C;Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
C;Accession: A49512
R;Laderman, K.A.; Asada, K.; Uemori, T.; Mukai, H.; Taguchi, Y.; Kato, I.; Anfinsen,
J. Biol. Chem. 268, 24402-24407, 1993
A;Title: alpha-Amylase from the hyperthermophilic archaeobacterium Pyrococcus furiosus
A;Reference number: A49512; MUID:94043280; PMID:8226990
A;Accession: A49512
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-649 <LAD>
A;Cross-references: GB:L22346; NID:g347939; PIDN:AAA72035.1; PID:g347940
C;Genetics:
A;Gene: amYA
A;Start codon: GTG
C;Function:
A;Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A;Pathway: glycogen/starch degradation
C;Superfamily: Dictyoglomus thermophilum amylase A
C;Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 7.5%; Score 140.5; DB 2; Length 649;
Best Local Similarity 22.8%; Pred. No. 0.01;
Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;
QY 25 EKAYIPVIETLIKEIP---FGLNITGYTLKFLP---KDIIDLKVGGIASDLIEIGTSY 78
||| | :||| || | : : | :||| : ||| : : ||| :

Db 29 EKCYPWFLETL--EYYPNMKVIAHTSGPLIEWLQDNRPYIDLLSLVKRGQVEIVVAGF 86

QY 79 THAILPLPLSRVEAQVQDRREVKE--ELFEVSPKGFWLPELAYDPIIPAILKDNNGEYEL 136

Db 87 YEPVLASIP---KEDRIEQIRLMKEWAKSIGFDARGVWLTQVWQPELVKTLKESGIDYV 143

QY 137 FADGEAMLFSAHLNSAIKPIKPLY-PHLIKAQRE-----KRFRYISYLLGLRELKRA 187

Db 144 IVD-----DYHFSAGLSKEELYWPYYTEDGGGEVIAVFPIDEKLR---YLIPFRPVDKV 194

QY 188 IKL---VFEGKVTLKAV--KDIEAVPVVAVNTAVMLGIGRLPLMPNPKKVASWI----- 236

Db 195 LEYLHSLIDGDESKVAVFDDGKFGIWPBTYEWVY-----EKGWLREFFDR 241

QY 237 ---EDKDNILLYGTDIE-----FIGYRDIAGYRMSVEGLLEVIDELNSCLP----- 281

Db 242 ISSDEKINMLYTELEKYKPRGLVYLPITASY-----FEM-----SEWSLPKQARLF 289

QY 282 ----SELKHS-----RELYLRTSSWAPDKSLRIWREDEGN---ARLNMLSYNMRGELAF 330

Db 290 VEFVNELKVGIFEKYRVFVRGGIW---KNF-FYKYPESNYMHKRMMLVSKLVRNN--- 341

QY 331 AENSARGWEPLPERLDRAFRAYND--WRG 359

Db 342 -----PEARKYLLRAQCNDAYWHG 360

RESULT 10

D71011

hypothetical protein PH1386 - Pyrococcus horikoshii

C;Species: Pyrococcus horikoshii

C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000

C;Accession: D71011

R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi DNA Res. 5, 55-76, 1998

A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic A;Reference number: A71000; MUID:98344137; PMID:9679194

A;Accession: D71011

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-560 <KAW>

A;Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30492.1; PID:g3257809

A;Experimental source: strain OT3

A;Note: this accession replaces an interim accession for a sequence replaced by GenBank C;Genetics:

A;Gene: PH1386

C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

Query Match 6.9%; Score 129.5; DB 2; Length 560;

Best Local Similarity 25.6%; Pred. No. 0.057;

Matches 60; Conservative 28; Mismatches 81; Indels 65; Gaps 12;

QY 14 EIPKSEIPKVIKAYIPVIETL-----IKEEIPFGLNITGYTLKFLPKDIIDL---VK 63

Db 71 EYIKREFEKYMERKLSMEEDLERFKDEKLREAINF---MIGYF-----KDVYSYWKSID 122

QY 64 GGIASDL-----IEIGTSYTHAILPLPLSR-VEAQVQDRREVKEELFEVSPKGF 114

Db 123 GNILGKFRELQDEGYVEVITSAATHGYLPLLDGEAIEAQLNGIKVYKFGKPRGIW 182

QY 115 LPELAYDPI-----IPAILKDNNGEYELFADGEAMLFSAHLNSAIKPIK 158

Db 183 LPECAIRPDGLWKSPTGEVKWRKGIEHFLKKGIEYFFVE-----SHLIDK-GPVSL 234

QY 159 LYPHLIKAQREKRFRYISYLLGLRELKRAIKLVFEGKVTLKAVKDIEAVPVVWA 212

Db 235 RYGNILPA-KTKRSTLRPYF-----LKNGIAVFARNRET-----GIQVWSA 274

RESULT 11

G83826

hypothetical protein BH1415 [imported] - Bacillus halodurans (strain C-125)

C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C;Accession: G83826

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a A;Reference number: A83650; MUID:20512582; PMID:11058132

A;Accession: G83826

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-923 <STO>

A;Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BAB05134.1; GSPDB:G A;Experimental source: strain C-125

C;Genetics:

A;Gene: BH1415

Query Match 6.8%; Score 127; DB 2; Length 923;

Best Local Similarity 23.8%; Pred. No. 0.18;

Matches 54; Conservative 31; Mismatches 70; Indels 72; Gaps 11;

QY 3 ALVPHGNLQYAEIPKSE-----IPKVIEKAYIPVIETLIKEEIPFGLNI----- 46

Db 7 SLVLHAHLPYVRHQEEDRLERWLFEAMSETYIPLLWALEKLPVKHAVTISFTPPVMEML 66

QY 47 -----TGY-----TLKFLPKD-----IIDLVKG-----G 65

Db 67 SDPLVQTRYLNHLENTEQLLKKEEKRTNDQRTQNLVQFYKQRYEKLKATFLQWDRNLLIG 126

QY 66 IASDLIE-----IIGTSYTHAILPLPLSR-VEAQVQDRREVKEELFEVSPKGFWLP 119

Db 127 FRS-LMENEQCTLMTSAATHAFPPYLKTKAIRAQVRHGIACFEQHFGKKPLGFWLPECA 185

QY 120 YDPIIPAILKDNNGEYELFADGEAMLF---SAHLNSAIKPIKPLY-PH 162

Db 186 FSPGVDRILFEEGIRYTFVDEHAVLTADPTPHKGSS----APIYSPH 228

RESULT 12

A75207

amylopullulanase PAB0122 - Pyrococcus abyssi (strain Orsay)

C;Species: Pyrococcus abyssi

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999

C;Accession: A75207

R;anonymous, Genoscope

A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s A;Reference number: A75001

A;Accession: A75207

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1362 <KAW>

A;Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49104.1; PID:el51 A;Experimental source: strain Orsay

C;Genetics:

A;Gene: apu; PAB0122

Query Match 6.7%; Score 126; DB 2; Length 1362;

Best Local Similarity 21.8%; Pred. No. 0.37;

Matches 83; Conservative 56; Mismatches 142; Indels 100; Gaps 16;

QY 32 IETLIKEEIPFGLNITGYTLKFLPKDIIDLVKGGIASDLIEIGTSYTHAILPLP 87

Db 236 VETVLKHQM-WLLNHT-----FEEHEKINLLG---NGNVEVTVPYTHPIGPI 286

QY 88 LSRVEAQVQDRREVKEELF---EVSPKGFWLPELAYDPIIPAILKDNNGEYELFADG 141

Db 287 YEDFDAQVKKANELYKEYLGAGKVTPKGGWAAESALNDKLTLEILAENGWKWMTDQLVLE 346

QY 142 AMLFSAHLNSAIKP-----IKPLYPHLIKA 166

Db 347 KLGVPKTIESYKPPVWAQFGDKKIYLFPRNHDLSDRVGFYAGMNQYDAVKNFVEELKI 406

QY 167 QREKRFRYISYLLGLRELKRAIKLVFECKVTLKAV-KDIEAVPVVAVNTAV-----ML 219

Db 407 QKQNYDGLVYVITLDGENPWEHPFDGKLFLEELYRQLEELQKKGLIRTVPSEYIEMF 466
QY 220 G--IGRLPLMNPVKVASWIEDKDNILLYGTDTIEFIGYRDIAGYRMSVEGLLEVIDELNSE 277
Db 467 GDKANKLTPKMKRLDFTTEDNVNALLKAKTL-----GELYDMVGVT----- 509
QY 278 LCLPSELKHSGRELYLRTSSWAPDKSLRIWREDEGNARLNMLSYNMRGELAFLAENS-D-A 336
Db 510 -----EMQWPESSWI-DGTLSTW---IGEPQENIAWYWLYLARKALFENKDNV 553
QY 337 RGWEPLPERRLDAFRAIYNDW 357
Db 554 KDNKAYEY---LFRAEGSDW 571

RESULT 13
AC1970
hypothetical protein alr1310 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC1970
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC1970
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-744 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073267.1; PID:gl17130657; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1310

Query Match 6.6%; Score 123.5; DB 2; Length 744;
Best Local Similarity 22.6%; Pred. No. 0.24;
Matches 79; Conservative 49; Mismatches 111; Indels 111; Gaps 19;
QY 58 SDLIEIIGTSYTHAILPL-----LPLSR-----VEAQVQDRREVKEELF 106
Db 208 SGQLEVTTPYTHPLPLADTNSGRVAVPNMALPESRFQWSEDIPIRLKAWELYTERF 267
QY 107 EVSPKGFWLPELAYDP-IIPAILKDNNGEYLFAD----GEAMLPFAHLNSAIKPIKP--L 159
Db 268 GOEPKGLWPSEQSVSPDILPYIIK-QGFQWICSDEAVLGTWTLKHFHRDGAGNVQOPELL 326
QY 160 Y-PHLIKAQ-----REKFRYISYLLGLRELKAIKLVFEGKVTLKAVKDIEAVPVW 210
Db 327 YRPYRLATPAGDLAIVFRDHR---LSDLIGF-----TYGAMPKQAAADL 368
QY 211 VAVNTAVMLGIGRLPLMNPKKVASWIEDKDNILLYGTD---IEFIGYRDIAGY-RMSVE 265
Db 369 VGHQAIAKMQREPSQEPWLVTIALDGENCWFEFYPQDGKPFLEAL-YQSLSNEPHIKLV 427
QY 266 GLEVIDELNSELCLPSELKHSGRELYLRTSSWAPDKSLRIWREDEGNARLNMLSYNMRG 325
Db 428 TVSEFIEEFPATATIPAEQLHSG-----SWV-DGSFTTWIGDP----- 464
QY 326 ELAFLAENSARGWEPLPERRL-----DAFRAIY-----NDW 357
Db 465 -----AKN---RAWDYLTARIMLANHPATEENNPEAWEAALYAAEGSDW 506

RESULT 14
C75120
hypothetical protein PAB1857 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: C75120
R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: C75120
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-602 <KAW>
A:Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49676.1; PID:g545
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1857
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

Query Match 6.3%; Score 119; DB 2; Length 602;
Best Local Similarity 26.8%; Pred. No. 0.39;
Matches 42; Conservative 20; Mismatches 59; Indels 36; Gaps 4;
QY 14 EIPKSEIPKVIKAYIPVETLIKEEIPFGLNITGYTLKFLPK-----DII 59
Db 71 EYIKAEFNRYMERKI-----KLMKEDLKADGKLRNAIEFMIKYFEDVYEWYWSKINGDII 125
QY 60 DLVKGGIASDLIEIIGTSYTHAILPLPLSLR-VEAQVQDRREVKEELFEVSPKGFWLPEL 118
Db 126 GRFKQLQDEGFEIITSAATHGYLPLLLGRDEAIDAQILTGIRVYKYGKFKPRGIWLPEC 185
QY 119 AYDPI-----IPAILKDNNGEYLFAD 139
Db 186 AYRPDGLWKSPTGEIKWRKGIHFLKKGYLEFFVE 222

RESULT 15
B69553
methanol dehydrogenase regulatory protein (moxr) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: B69553
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: B69553
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-324 <KLE>
A:Cross-references: GB:AE01108; GB:AE000782; NID:g2689431; PIDN:AAB91247.1; PID:g265
C:Superfamily: methanol dehydrogenase regulatory protein

Query Match 6.1%; Score 115.5; DB 1; Length 324;
Best Local Similarity 21.1%; Pred. No. 0.3;
Matches 69; Conservative 62; Mismatches 121; Indels 75; Gaps 14;
QY 1 MRALVFHGNLQYAEIP---KSEIPKVIKAYIPVETLIKEEIPFGLNITGYTLKFLPKD 57
Db 35 LAAALTNGNILFEDYPGLGKTLAKVFARV-----IGADYRRVQFTPDLLPSD 82
QY 58 IIDLVKGGI---ASDLIEII-GTSYTHAIL-----PLLPLSRVEAQVQDRREVKEEL 105
Db 83 II-----GVKIWRGDRFVKGPIFTNVLLADEINRSPPKTQAALLEAMEEKQITVEGT 137
QY 106 FEVSPKGFWLPELAYDPIIPAILKDNNGEYLFADGEAMLFSAHLSAIPKPIKPLYPHLIK 165
Db 138 FSLMPFFVL--ATQNPI-----EQEGTYPLPEAQMDREMLRMPPGYPESIE 182
QY 166 AQREKRRFYISY-----LLGLRELKAIKLVFEGKVTLKAVKDIEAVPVW 211
Db 183 EEMEILRRRISWRKDDPTEDVEPVVSLETRRIQDAVEAVYVDKSLKYLSELVRA---T 239
QY 212 AVNTAVMLGI---GRLPLMNPKKVASWIEDKDNIL---LYGTDIEFIGYRDIAGYRMSVE 265

Db 240 REHELVELGSSPRGGLALLKLARALAVMDGRDFVIPDDVKRVAVEALAHVRVILKFEYAVE 299

Qy 266 GLL--EVIDELNSELCLPSELKHSGRE 290

Db 300 GLRAEEVVEEILNSVRP---KYEAE 323

Search completed: November 18, 2002, 07:25:58
Job time : 27 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2002, 07:26:04 ; Search time 9 Seconds
(without alignments)
609.129 Million cell updates/sec

Title: US-09-619-032-4
Perfect score: 1880
Sequence: 1 MRALVFHGNLQYAEIPKSEI.....RRLDAFRAIYNDRWGENGEP 364

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1868	99.4	364	9	US-10-112-418-4 Sequence 4, Appli
2	1868	99.4	364	9	US-10-114-083-4 Sequence 4, Appli
3	1868	99.4	364	10	US-09-886-400-4 Sequence 4, Appli
4	1868	99.4	364	12	US-10-112-357-4 Sequence 4, Appli
5	1868	99.4	364	12	US-10-114-403-4 Sequence 4, Appli
6	1868	99.4	364	12	US-10-116-606-4 Sequence 4, Appli
7	1868	99.4	364	12	US-10-112-331-4 Sequence 4, Appli
8	1868	99.4	364	12	US-10-112-377-4 Sequence 4, Appli
9	1868	99.4	364	12	US-10-116-581-4 Sequence 4, Appli
10	1868	99.4	364	12	US-10-112-442-4 Sequence 4, Appli
11	103	5.5	1073	10	US-09-815-242-12361 Sequence 12361, A
12	103	5.5	1147	10	US-09-815-242-5468 Sequence 5468, Ap
13	91	4.8	813	10	US-09-732-180-7 Sequence 7, Appli
14	91	4.8	844	10	US-09-732-180-2 Sequence 2, Appli
15	89.5	4.8	1668	10	US-09-815-242-5654 Sequence 5654, Ap
16	89.5	4.8	2397	10	US-09-815-242-12265 Sequence 12265, A
17	88.5	4.7	764	10	US-09-828-466-4 Sequence 4, Appli
18	88.5	4.7	993	10	US-09-894-998-50 Sequence 50, Appli
19	88.5	4.7	1037	10	US-09-894-998-54 Sequence 54, Appli

20	88.5	4.7	1113	10	US-09-894-998-51	Sequence 51, Appli
21	87.5	4.7	760	10	US-09-764-367A-9	Sequence 9, Appli
22	87	4.6	657	10	US-09-950-368-1	Sequence 1, Appli
23	86.5	4.6	764	10	US-09-828-466-5	Sequence 5, Appli
24	86.5	4.6	1357	10	US-09-815-242-11997	Sequence 11997, A
25	85.5	4.5	511	9	US-10-121-032-21	Sequence 21, Appli
26	83.5	4.4	354	10	US-09-815-242-12884	Sequence 12884, A
27	83.5	4.4	478	10	US-09-815-242-5719	Sequence 5719, Ap
28	83.5	4.4	478	10	US-09-815-242-12480	Sequence 12480, A
29	82.5	4.4	237	10	US-09-815-242-5666	Sequence 5666, Ap
30	82.5	4.4	258	10	US-09-815-242-12207	Sequence 12207, A
31	82.5	4.4	788	10	US-09-391-340-6	Sequence 6, Appli
32	82.5	4.4	788	10	US-09-948-369-6	Sequence 6, Appli
33	82	4.4	322	10	US-09-838-561-11	Sequence 11, Appli
34	82	4.4	322	10	US-09-816-760-11	Sequence 11, Appli
35	81.5	4.3	354	10	US-09-815-242-12643	Sequence 12643, A
36	81.5	4.3	804	10	US-09-815-242-4982	Sequence 4982, Ap
37	81.5	4.3	806	10	US-09-815-242-11622	Sequence 11622, A
38	81	4.3	303	10	US-09-838-561-13	Sequence 13, Appli
39	81	4.3	303	10	US-09-816-760-13	Sequence 13, Appli
40	80.5	4.3	1251	10	US-09-778-927A-58	Sequence 58, Appli
41	80.5	4.3	1602	10	US-09-778-927A-59	Sequence 59, Appli
42	80	4.3	224	10	US-09-764-853-534	Sequence 534, App
43	80	4.3	535	10	US-09-815-242-13801	Sequence 13801, A
44	80	4.3	541	12	US-10-047-676A-11	Sequence 11, Appli
45	80	4.3	679	10	US-09-815-242-13418	Sequence 13418, A

ALIGNMENTS

RESULT 1
US-10-112-418-4
; Sequence 4, Application US/10112418
; Patent No. US20020155486A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Murphy, Dennis
; APPLICANT: Ried, John
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: DIVER1120-4
; CURRENT APPLICATION NUMBER: US/10/112,418
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/886,400
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 09/619,032
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/407,806
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 08/613,220
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Thermococcus alcaliphilus
US-10-112-418-4

Query Match 99.4%; Score 1868; DB 9; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVETLKEIPFGLNITGYTLKFLPKDIID 60
:|||||
Db 1 LRALVFHGNLQYAEIPKSEIPKVIKAYIPVETLKEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQRDRVKEELFEVSPKGFWLPDELAY 120
:|||||
Db 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQRDRVKEELFEVSPKGFWLPDELAY 120
QY 121 DPTIIPAILKDNNGEYLFADGEAMLFSAHLNSAIKPIKPLYPHLIKAQREKFRFYISYLLG 180

|||||
Db 121 DPIIPAILKONGEYELFADGEAMLFSAHLSNAIKPIKPLYPHLIKAQREKFRYISYLLG 180
QY 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCKKVASWIEDKD 240
Db 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCKKVASWIEDKD 240
QY 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
Db 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
QY 361 NGEF 364
Db 361 NGEF 364

RESULT 2

US-10-114-083-4
; Sequence 4, Application US/10114083
; Patent No. US20020160464A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Murphy, Dennis
; APPLICANT: Ried, John
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND
; FILE REFERENCE: DIVER1120-4
; CURRENT APPLICATION NUMBER: US/10/114,083
; PRIOR FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 09/886,400
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 09/619,032
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/407,806
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 08/613,220
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Thermococcus alcaliphilus
US-10-114-083-4

Query Match 99.4%; Score 1868; DB 9; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKEEIPFGLNITGYTLKFLPKDIID 60
Db 1 LRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPPELAY 120
Db 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPPELAY 120
QY 121 DPIIPAILKONGEYELFADGEAMLFSAHLSNAIKPIKPLYPHLIKAQREKFRYISYLLG 180
Db 121 DPIIPAILKONGEYELFADGEAMLFSAHLSNAIKPIKPLYPHLIKAQREKFRYISYLLG 180
QY 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCKKVASWIEDKD 240
Db 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCKKVASWIEDKD 240
QY 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300

QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
Db 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
QY 361 NGEF 364
Db 361 NGEF 364

RESULT 3

US-09-886-400-4
; Sequence 4, Application US/09886400
; Patent No. US20020045226A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Murphy, Dennis
; APPLICANT: Ried, John
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE
; FILE REFERENCE: DIVER1120-4
; CURRENT APPLICATION NUMBER: US/09/886,400
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 09/619,032
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/407,806
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 08/613,220
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Thermococcus alcaliphilus
US-09-886-400-4

Query Match 99.4%; Score 1868; DB 10; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKEEIPFGLNITGYTLKFLPKDIID 60
Db 1 LRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPPELAY 120
Db 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPPELAY 120
QY 121 DPIIPAILKONGEYELFADGEAMLFSAHLSNAIKPIKPLYPHLIKAQREKFRYISYLLG 180
Db 121 DPIIPAILKONGEYELFADGEAMLFSAHLSNAIKPIKPLYPHLIKAQREKFRYISYLLG 180
QY 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCKKVASWIEDKD 240
Db 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPCKKVASWIEDKD 240
QY 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
Db 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
QY 361 NGEF 364
Db 361 NGEF 364

RESULT 4

US-10-112-357-4
; Sequence 4, Application US/10112357
; Patent No. US20020115099A1
; GENERAL INFORMATION:

APPLICANT: DIVERSA CORPORATION
APPLICANT: Murphy, Dennis
APPLICANT: Ried, John
TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE THEREOF
FILE REFERENCE: DIVER1120-4
CURRENT APPLICATION NUMBER: US/10/112,357
CURRENT FILING DATE: 2002-03-29
PRIOR APPLICATION NUMBER: 09/886,400
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 09/619,032
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/407,806
PRIOR FILING DATE: 1999-09-20
PRIOR APPLICATION NUMBER: 08/613,220
PRIOR FILING DATE: 1996-03-08
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 364
TYPE: PRT
ORGANISM: Thermococcus alcaliphilus
US-10-112-357-4

Query Match 99.4%; Score 1868; DB 12; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKEIPFGLNITGYTLKFLPKDIID 60
:|||||
Db 1 LRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKEIPFGLNITGYTLKFLPKDIID 60
:|||||

QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPELAY 120
:|||||
Db 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPELAY 120
:|||||

QY 121 DPIIPAILKDNQYGYELFADGEAMLFSAHLSAIPKPIKPLYPHLIKAQREKFRFYISYLLG 180
:|||||
Db 121 DPIIPAILKDNQYGYELFADGEAMLFSAHLSAIPKPIKPLYPHLIKAQREKFRFYISYLLG 180
:|||||

QY 181 LRELKAIKLVFEGKVTCLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
:|||||
Db 181 LRELKAIKLVFEGKVTCLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
:|||||

QY 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
:|||||
Db 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
:|||||

QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
:|||||
Db 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
:|||||

QY 361 NGEF 364
:|||||
Db 361 NGEF 364

RESULT 5
US-10-114-403-4
Sequence 4, Application US/10114403
Patent No. US20020115100A1
GENERAL INFORMATION:
APPLICANT: DIVERSA CORPORATION
APPLICANT: Murphy, Dennis
APPLICANT: Ried, John
TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE THEREOF
FILE REFERENCE: DIVER1120-4
CURRENT APPLICATION NUMBER: US/10/114,403
CURRENT FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: 09/886,400
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 09/619,032
PRIOR FILING DATE: 2000-07-19

PRIOR APPLICATION NUMBER: 09/407,806
PRIOR FILING DATE: 1999-09-20
PRIOR APPLICATION NUMBER: 08/613,220
PRIOR FILING DATE: 1996-03-08
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 364
TYPE: PRT
ORGANISM: Thermococcus alcaliphilus
US-10-114-403-4

Query Match 99.4%; Score 1868; DB 12; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKEIPFGLNITGYTLKFLPKDIID 60
:|||||
Db 1 LRALVFHGNLQYAEIPKSEIPKVIKAYIPVIETLIKEIPFGLNITGYTLKFLPKDIID 60
:|||||

QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPELAY 120
:|||||
Db 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPELAY 120
:|||||

QY 121 DPIIPAILKDNQYGYELFADGEAMLFSAHLSAIPKPIKPLYPHLIKAQREKFRFYISYLLG 180
:|||||
Db 121 DPIIPAILKDNQYGYELFADGEAMLFSAHLSAIPKPIKPLYPHLIKAQREKFRFYISYLLG 180
:|||||

QY 181 LRELKAIKLVFEGKVTCLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
:|||||
Db 181 LRELKAIKLVFEGKVTCLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
:|||||

QY 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
:|||||
Db 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
:|||||

QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
:|||||
Db 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSDARGWEPLPERRLDAFRAIYNDWRGE 360
:|||||

QY 361 NGEF 364
:|||||
Db 361 NGEF 364

RESULT 6
US-10-116-606-4
Sequence 4, Application US/10116606
Patent No. US20020119515A1
GENERAL INFORMATION:
APPLICANT: DIVERSA CORPORATION
APPLICANT: Murphy, Dennis
APPLICANT: Ried, John
TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE
FILE REFERENCE: DIVER1120-4
CURRENT APPLICATION NUMBER: US/10/116,606
CURRENT FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 09/886,400
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 09/619,032
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/407,806
PRIOR FILING DATE: 1999-09-20
PRIOR APPLICATION NUMBER: 08/613,220
PRIOR FILING DATE: 1996-03-08
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 364
TYPE: PRT
ORGANISM: Thermococcus alcaliphilus
US-10-116-606-4

Query Match 99.4%; Score 1868; DB 12; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
Db :|||||
1 LRALVFHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPDELAY 120
Db :|||||
61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPDELAY 120
QY 121 DPTIPAILKONGYELFADGEAMLFSAHLSAIPKIPKLYPHLIKAQREKFRFYISYLLG 180
Db :|||||
121 DPTIPAILKONGYELFADGEAMLFSAHLSAIPKIPKLYPHLIKAQREKFRFYISYLLG 180
QY 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
Db :|||||
181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
QY 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db :|||||
241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSADARGWEPLPERRLDAFRAIYNDWRGE 360
Db :|||||
301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSADARGWEPLPERRLDAFRAIYNDWRGE 360
QY 361 NGEF 364
Db :|||||
361 NGEF 364

RESULT 7
US-10-112-331-4
; Sequence 4, Application US/10112331
; Patent No. US20020119550A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Murphy, Dennis
; APPLICANT: Ried, John
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE TH
; FILE REFERENCE: DIVER1120-4
; CURRENT APPLICATION NUMBER: US/10/112,331
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US/09/886,400
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 09/619,032
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/407,806
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 08/613,220
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Thermococcus alcaliphilus
US-10-112-331-4

Query Match 99.4%; Score 1868; DB 12; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
Db :|||||
1 LRALVFHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPDELAY 120
Db :|||||
61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPDELAY 120

QY 121 DPTIPAILKONGYELFADGEAMLFSAHLSAIPKIPKLYPHLIKAQREKFRFYISYLLG 180
Db :|||||
121 DPTIPAILKONGYELFADGEAMLFSAHLSAIPKIPKLYPHLIKAQREKFRFYISYLLG 180
QY 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
Db :|||||
181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
QY 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db :|||||
241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
QY 301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSADARGWEPLPERRLDAFRAIYNDWRGE 360
Db :|||||
301 DKSLRIWREDEGNARLNMLSYNMRGELAFLAENSADARGWEPLPERRLDAFRAIYNDWRGE 360
QY 361 NGEF 364
Db :|||||
361 NGEF 364

RESULT 8
US-10-112-377-4
; Sequence 4, Application US/10112377
; Patent No. US20020120108A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Murphy, Dennis
; APPLICANT: Ried, John
; TITLE OF INVENTION: ENZYMES HAVING ALPHA-GALACTOSIDASE ACTIVITY AND METHODS OF USE
; FILE REFERENCE: DIVER1120-4
; CURRENT APPLICATION NUMBER: US/10/112,377
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/886,400
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 09/619,032
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/407,806
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 08/613,220
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Thermococcus alcaliphilus
US-10-112-377-4

Query Match 99.4%; Score 1868; DB 12; Length 364;
Best Local Similarity 99.2%; Pred. No. 1.6e-164;
Matches 361; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRALVFHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
Db :|||||
1 LRALVFHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPDELAY 120
Db :|||||
61 LVKGGIASDLIEIIGTSYTHAILPLPLSRVEAQVQDRREVKEELFEVSPKGFWLPDELAY 120
QY 121 DPTIPAILKONGYELFADGEAMLFSAHLSAIPKIPKLYPHLIKAQREKFRFYISYLLG 180
Db :|||||
121 DPTIPAILKONGYELFADGEAMLFSAHLSAIPKIPKLYPHLIKAQREKFRFYISYLLG 180
QY 181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
Db :|||||
181 LRELKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKD 240
QY 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300
Db :|||||
241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAP 300


```
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12361
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12361
```

```
Query Match          5.5%; Score 103; DB 10; Length 1073;
Best Local Similarity 21.2%; Pred. No. 0.28;
Matches 58; Conservative 48; Mismatches 90; Indels 78; Gaps 15;

QY 107 EVSPKGFWLPPELAYDPIIPAILKDNNGEYELFADGEAMLFSAHLSAIPKIPLYPHLIKA 166
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 309 EISP-----YYDSL---LVKLSTHAISFKQAEKMWRSRLREMRIRGVKTNIPFLINV 357
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 167 QREKRF-----RYISYLLGLRELKAI-----KLVFEGKVTLKAVKDIE----- 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 358 MNKKFTSGDYTTKFIETPELFDIQPSLDRGTKTLEYIGNVTINGFPNVEKRPKPDYEL 417
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 206 -AVPVWVAVNTAVMLGIGR-LPLMNPKKVASWIEDKDNILLYGTDIEF-IGYRDIAGYRM 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 ASIPTVSSSKIASFGTKQLLDEVGPKGVAEWVKQDDVLL--TDTFRDAHQSLLATRV 475
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 263 SVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAPDK-SLRWREDEGNARLNMLSY 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 476 RTKDMINIAS-----KTADVFKDGFSLMW-----GGATFD-VAY 509

QY 322 NMRGELAFLAENSDARGWEPLPERRLLDAFRAIYN 355
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 510 N-----FLKENP-----WERLERLR-----KAIPN 529
```

```
RESULT 12
US-09-815-242-5468
; Sequence 5468, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5468
; LENGTH: 1147
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5468
```

```
Query Match          5.5%; Score 103; DB 10; Length 1147;
Best Local Similarity 21.2%; Pred. No. 0.31;
Matches 58; Conservative 48; Mismatches 90; Indels 78; Gaps 15;

QY 107 EVSPKGFWLPPELAYDPIIPAILKDNNGEYELFADGEAMLFSAHLSAIPKIPLYPHLIKA 166
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 385 EISP-----YYDSL---LVKLSTHAISFKQAEKMWRSRLREMRIRGVKTNIPFLINV 433
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 167 QREKRF-----RYISYLLGLRELKAI-----KLVFEGKVTLKAVKDIE----- 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 434 MNKKFTSGDYTTKFIETPELFDIQPSLDRGTKTLEYIGNVTINGFPNVEKRPKPDYEL 493
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 206 -AVPVWVAVNTAVMLGIGR-LPLMNPKKVASWIEDKDNILLYGTDIEF-IGYRDIAGYRM 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 494 ASIPTVSSSKIASFGTKQLLDEVGPKGVAEWVKQDDVLL--TDTFRDAHQSLLATRV 551
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 263 SVEGLLEVIDELNSELCLPSELKHSGRELYLRTSSWAPDK-SLRWREDEGNARLNMLSY 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 552 RTKDMINIAS-----KTADVFKDGFSLMW-----GGATFD-VAY 585

QY 322 NMRGELAFLAENSDARGWEPLPERRLLDAFRAIYN 355
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 586 N-----FLKENP-----WERLERLR-----KAIPN 605
```

```
RESULT 13
US-09-732-180-7
; Sequence 7, Application US/09732180
; Patent No. US20020137702A1
; GENERAL INFORMATION:
; APPLICANT: Arkowitz, Robert A
; APPLICANT: Nern, Peter MA
; TITLE OF INVENTION: Yeast receptor
; FILE REFERENCE: DY025.001AUS
; CURRENT APPLICATION NUMBER: US/09/732,180
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,699
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-732-180-7
```

```
Query Match          4.8%; Score 91; DB 10; Length 813;
Best Local Similarity 22.1%; Pred. No. 2.4;
Matches 81; Conservative 56; Mismatches 133; Indels 96; Gaps 19;

QY 25 EKAYIPVIFTLIK--EEIPFGLNITGYTLKFL---PKDIIDLVK---GGIASDL-----I 71
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 ERKYVQDLELMCKYRQDLIEAENLSSEQIHLLFPNLNEIIDFQRRFLNGLECNINPIRY 334
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 EIIGTSYTHAIL-----PLLPLSRVEAQVQDRDREVKELFEVSPKGFWLPPELAYDP 122
```

Db 335 QRIQSVFIHASLGFNAYEPWTIGLTALDILINKEAANKSSSLDP-GFELQSYILKP 393
QY 123 I-----IPAILKD---NGEYLFADGEAMLFSAHLNS---AIKPIKPLYPHLIKAQREKR 171
Db 394 IORLCKYPLLLKELIKTSPEYSKQDPHGSSSTSFNELLVAKTAMKELANQVNEAQR--R 451
QY 172 FRYISYLLGLRE-----LRKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIG 222
Db 452 AENIEHLEKLERVGNWRGNLDAQGELLFHGQV---GVKDAE----- 491
QY 223 RLPLMNPKKVASWIEDKDNILLYGTDIEFIGYRD-----IAGYRMSVEGL 267
Db 492 -----NEKEYVAYLFK--IVFFTEIDDTKKSDKQEKSKFSTRKRSTSSNLSSSTTNL 544
QY 268 LEVIDELNSELCLPSELKHSGR---ELYLRTSSWAPDKSLRI-W--REDEGNARLNMLS 320
Db 545 LESINNSRKDNTLPLELK--GRVYISEIYNISAPNTPGSTLIISWSGRKESGSFTLRYRS 602
QY 321 YNMERGE 326
Db 603 EARNQ 608

RESULT 14
US-09-732-180-2
; Sequence 2, Application US/09732180
; Patent No. US20020137702A1
; GENERAL INFORMATION:
; APPLICANT: Arkowitz, Robert A
; APPLICANT: Nern, Peter MA
; TITLE OF INVENTION: Yeast receptor
; FILE REFERENCE: DYOU25.001AUS
; CURRENT APPLICATION NUMBER: US/09/732,180
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,699
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 844
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-732-180-2

Query Match 4.8%; Score 91; DB 10; Length 844;
Best Local Similarity 22.1%; Pred. No. 2.6;
Matches 81; Conservative 56; Mismatches 133; Indels 96; Gaps 19;
QY 25 EKAYIPVIETLIK--EPIPFGLNITGYTLKFL---PKDIIDLK---GGIASDL-----I 71
Db 289 ERKYVDLELMCKYRQDLIEAENLSSEQIHLFPNLNEIIDFQRRFLNGLECNINVPY 348
QY 72 EIIGTSYTHAIL-----PLPLSRVBAQVQORDREVKEELFEVSPKGFWLPDELAYDP 122
Db 349 QRIQSVFIHASLGFNAYEPWTIGLTALDILINKEAANKSSSLDP-GFELQSYILKP 407
QY 123 I-----IPAILKD---NGEYLFADGEAMLFSAHLNS---AIKPIKPLYPHLIKAQREKR 171
Db 408 IORLCKYPLLLKELIKTSPEYSKQDPHGSSSTSFNELLVAKTAMKELANQVNEAQR--R 465
QY 172 FRYISYLLGLRE-----LRKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMLGIG 222
Db 466 AENIEHLEKLERVGNWRGNLDAQGELLFHGQV---GVKDAE----- 505
QY 223 RLPLMNPKKVASWIEDKDNILLYGTDIEFIGYRD-----IAGYRMSVEGL 267
Db 506 -----NEKEYVAYLFK--IVFFTEIDDTKKSDKQEKSKFSTRKRSTSSNLSSSTTNL 558
QY 268 LEVIDELNSELCLPSELKHSGR---ELYLRTSSWAPDKSLRI-W--REDEGNARLNMLS 320
Db 559 LESINNSRKDNTLPLELK--GRVYISEIYNISAPNTPGSTLIISWSGRKESGSFTLRYRS 616

QY 321 YNMERGE 326
Db 617 EARNQ 622
RESULT 15
US-09-815-242-5654
; Sequence 5654, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5654
; LENGTH: 1668
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5654

Query Match 4.8%; Score 89.5; DB 10; Length 1668;
Best Local Similarity 19.6%; Pred. No. 9.2;
Matches 67; Conservative 68; Mismatches 128; Indels 79; Gaps 15;
QY 17 KSEIPKVIKAYIPVIETLIKKEIIPFGLN-----ITGYTLK 52
Db 513 KSQRLMTLPEYMIPIV-NFMHIEQIPITINGKLDKKALPIMDYVDTDAYVAPSTDTEHLLC 571
QY 53 FLPKDIIDLKGGIASDLIEIIGTSYTHAILPLPLSRVBAQVQORDREV-----KEELFE 107
Db 572 QIFADILHVNQVGIHDNFFELGG---HSLKATLVVNRIEASTGKRQIGDILLQKPTVFE 627
QY 108 VSPKGFWLPDELAYDPIIPAILKNDNGEYELFADGEAMLF-----SAHLNSAIKPIKPLYPHL 163
Db 628 LAQATAKVOEQNYEVIPEAIVKD---DYVLSSAQKRMVLLWKSNNHKDTVYN-----VPFL 679
QY 164 IKAQREKRF-----RYISYLLGLRELKRAIKLVFEGKVTLKAVKDIEAVPVVAVNT----- 215
Db 680 WRLSSELNVAQLRQAVQHILARHEILRTQYIVVDDEVQRIVADV--VADFEVNTHTFTD 737
QY 216 ---AVMLGIGRLPLMNPKKV-ASWIEDKDNILLYGTDIEFIGYRDIAGYRMSVEGLLEVI 271
Db 738 EQEIMRQFVAPFNLEKPSQIRVYIRSPHAYL-----FIDTHHIINDGMS---NIQLM 788
QY 272 DELNS---ELCLPSELKHSGRRELYLRTSSWAPDKSLRIWRE 309
Db 789 NDLNALYQHKLLPLKIQ-----YKDYSEWMSHRDMTKHRQ 824

Search completed: November 18, 2002, 07:30:03
Job time : 11 secs
